

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 15:46:52 ; Search time 10065 Seconds  
(without alignments)  
11186.962 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agcttctgcgcagaaagg.....atctgaagctcactattcag 2381

Scoring table: OLIGO-NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	437	18.4	2847	5	AY254202	Gallus ga
2	85	3.6	923	5	BX932340	Gallus ga
3	81	3.4	899	5	BX930518	Gallus ga
4	78	3.3	896	5	BX933597	Gallus ga
5	27	1.1	198504	2	AC121429	Rattus no
6	27	1.1	240387	2	AC107353	Rattus no
7	26	1.1	115199	9	AC003695	Homo sapi
8	26	1.1	133455	9	AC009451	Homo sapi
9	26	1.1	179668	2	CR626936	Danio rer
10	26	1.1	199534	2	AC025959	Homo sapi
11	26	1.1	205195	2	AC025963	Homo sapi
12	25	1.0	132948	9	HS349A12	Human DNA
13	25	1.0	141876	2	AC141717	Apis mell
14	25	1.0	166622	2	AC068973	Homo sapi
15	25	1.0	167118	9	AC023385	Homo sapi
16	25	1.0	168711	2	AC087201	Homo sapi
17	25	1.0	168917	2	CR354390	Danio rer
18	25	1.0	223843	10	AC102595	Mus muscu
19	25	1.0	238113	2	AC094216	Rattus no

20	1.0	242282	5	AL953891	Zebrafish
21	1.0	1160	6	BD276057	48 Human
22	1.0	4911	8	AF379855	Carica pa
23	1.0	6130	6	AX251480	Sequence
24	1.0	6130	6	AX252113	Sequence
25	1.0	7133	6	AX251142	Sequence
26	1.0	14810	2	AC014701	Drosophila
27	1.0	22901	6	CQ598715	Sequence
28	1.0	42242	1	AV372455	Unculture
29	1.0	52154	9	AC090643	Homo sapi
30	1.0	89017	2	AC108387	Pan trogl
31	1.0	110000	8	CR382125	Continuation (8 of
32	1.0	120720	2	AC146498	Gallus ga
33	1.0	139017	2	AC150803	Macaca mu
34	1.0	145704	3	AE002656	Drosophil
35	1.0	154027	2	BX942840	Danio rer
36	1.0	157805	10	AL772183	Mouse DNA
37	1.0	168059	3	AC010027	Drosophil
38	1.0	168283	2	AC027633	Homo sapi
39	1.0	172942	9	AC123537	Macaca mu
40	1.0	173175	2	BX927368	Danio rer
41	1.0	174408	2	AC150303	Papio anu
42	1.0	178161	2	AC150825	Callithri
43	1.0	180056	2	CR354613	Danio rer
44	1.0	182982	9	AL355522	Human DNA
45	1.0	183881	2	AC138725	Cercopith

#### ALIGNMENTS

RESULT 1	AY254202	2847 bp	DNA	linear	VRT 22-APR-2003
LOCUS	Gallus gallus intestinal fatty acid-binding protein gene, complete cds.				
DEFINITION	AY254202				
ACCESSION	AY254202.1	GI:30060211			
VERSION					
KEYWORDS	Gallus gallus (chicken)				
SOURCE	Gallus gallus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (bases 1 to 2847)				
AUTHORS	Wang,Q., Li,H., Wang,Y. and Zhao,J.				
TITLE	Cloning and characterization of chicken I-FABP gene				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2847)				
AUTHORS	Wang,Q., Li,H., Wang,Y. and Zhao,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-MAR-2003) Animal Science & Technology, Northeast Agricultural University, 59 Mucai Street, Harbin, Heilongjiang 150030, China				
FEATURES	Location/Qualifiers				
source	1..2847				
	/organism="Gallus gallus"				
	/mol_type="Genomic DNA"				
	/db_xref="taxon:9031"				
	/tissue_type="blood"				
	/notes="Breed: Arber Acres broiler"				
mRNA	join(<21..87,720..892,1465..1572,2416..>2466)				
	/product="intestinal fatty acid-binding protein"				
CDS	join(21..87,720..892,1465..1572,2416..2466)				
	/notes="FABP"				
	/codon_start=1				
	/product="intestinal fatty acid-binding protein"				
	/protein_id="AAP13101.1"				
	/db_xref="GI:30060212"				
	/translation="MAFGWTGKIEKNYEKFMAGVNVNKKKLGAGHNDKLKTIQDDGNKFLVKSSNFRTIDIEFTLGVSPYSLADGTSLGSWNLEGNKLVGTFTRKDNGKVLTAIRBIVGSELITQTVYVEGVAKRFKKE"				
ORIGIN					

```
Query Match      18.4%; Score 437; DB 5; Length 2847;
Best Local Similarity 100.0%; Pred. No. 6.1e-212;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1617 AGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGAAA 1676
Db 11 AGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGAAA 70

QY 1677 ATTATGGAAGCAATGGGTAAGCCCTACTTTTGAATGCCCTTCTAAAGCAGGATACCA 1736
Db 71 ATTATGGAAGCAATGGGTAAGCCCTACTTTTGAATGCCCTTCTAAAGCAGGATACCA 130

QY 1737 CTACGGCGGAATACAAACTTAAGCTGTTCATGAACCTACCATCTGGCTAACTGTCCTTTG 1796
Db 131 CTACGGCGGAATACAAACTTAAGCTGTTCATGAACCTACCATCTGGCTAACTGTCCTTTG 190

QY 1797 TTGCTGCTATTTTGGCCCTTGACATTTGCCCTGACATTTATTTGAAAGACTCTATAGAG 1856
Db 191 TTGCTGCTATTTTGGCCCTTGACATTTGCCCTGACATTTATTTGAAAGACTCTATAGAG 250

QY 1857 GGGATACAGAGAAACAACTCTGATTTTATTTGATTCGATTCGCATTAATCTTATGCAATTT 1916
Db 251 GGGATACAGAGAAACAACTCTGATTTTATTTGATTCGATTCGCATTAATCTTATGCAATTT 310

QY 1917 AGCTAATTCAGTAGAGGCATTCAGACGAAATTTAAATAGATTAATCTTATGAGGATAT 1976
Db 311 AGCTAATTCAGTAGAGGCATTCAGACGAAATTTAAATAGATTAATCTTATGAGGATAT 370

QY 1977 TATTTGTATAGACTGTTTGAATAATACACAGAGGAAATTCGCTGCTCCAGTTTGTG 2036
Db 371 TATTTGTATAGACTGTTTGAATAATACACAGAGGAAATTCGCTGCTCCAGTTTGTG 430

QY 2037 CAGAACACACATGATTT 2053
Db 431 CAGAACACACATGATTT 447

RESULT 2
BX932340
LOCUS      923 bp      mRNA      linear      VRT 30-MAR-2004
DEFINITION Gallus gallus finished cDNA, clone CHEST355022.
ACCESSION  BX932340
VERSION     BX932340.2 GI:46017395
KEYWORDS   Gallus gallus (chicken)
SOURCE     Gallus gallus
ORGANISM   Gallus gallus
REFERENCE  Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
AUTHORS    Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
            Tickle,C. and Wilson,S.A.
            Direct Submission
            Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: chickestbms.unist.ac.uk
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
            sequencing project.
            This sequence is from the
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
            from a library constructed by Elizabeth Bosch. cDNA was prepared
            from RNA extracted from small intestine, normalised, and poly
            A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.
            Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host:
            Escherichia coli DH10B.
            Location/Qualifiers
                1..893
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Compton Line 151"
                /db_xref="taxon:9031"
                /clone="CHEST355022"
                /dev_stage="adult"

TITLE
JOURNAL
COMMENT

FEATURES
source
    1..893
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Compton Line 151"
    /db_xref="taxon:9031"
    /clone="CHEST355022"
    /dev_stage="adult"

ORIGIN
Query Match      3.4%; Score 81; DB 5; Length 899;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1614 TACAGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGA 1673
Db 11 TACAGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGA 70

QY 1674 AAAATTCATGGAAGCAATGGG 1694
Db 71 AAAATTCATGGAAGCAATGGG 91
```

```
ORIGIN
Query Match      3.6%; Score 85; DB 5; Length 923;
Best Local Similarity 100.0%; Pred. No. 1.4e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1610 TGCTACACAGACAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACT 1669
Db 1 TGCTACACAGACAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACT 60

QY 1670 ATGAAAAATTCATGGAAGCAATGGG 1694
Db 61 ATGAAAAATTCATGGAAGCAATGGG 85

RESULT 3
BX930518
LOCUS      899 bp      mRNA      linear      VRT 02-FEB-2004
DEFINITION Gallus gallus finished cDNA, clone CHEST640b17.
ACCESSION  BX930518
VERSION     BX930518.1 GI:41631046
KEYWORDS   Gallus gallus (chicken)
SOURCE     Gallus gallus
ORGANISM   Gallus gallus
REFERENCE  Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
AUTHORS    Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
            Tickle,C. and Wilson,S.A.
            Direct Submission
            Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: chickestbms.unist.ac.uk
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
            sequencing project.
            This sequence is from the
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
            from a library constructed by Elizabeth Bosch. cDNA was prepared
            from RNA extracted from small intestine, normalised, and poly
            A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.
            Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host:
            Escherichia coli DH10B.
            Location/Qualifiers
                1..899
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Compton Line 151"
                /db_xref="taxon:9031"
                /clone="CHEST640b17"
                /dev_stage="adult"

TITLE
JOURNAL
COMMENT

FEATURES
source
    1..899
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Compton Line 151"
    /db_xref="taxon:9031"
    /clone="CHEST640b17"
    /dev_stage="adult"

ORIGIN
Query Match      3.4%; Score 81; DB 5; Length 899;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1614 TACAGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGA 1673
Db 11 TACAGACAGAAAGATGGCATTTTAACGGTACTTGGAAATAGAGAAATAGAACTATGA 70

QY 1674 AAAATTCATGGAAGCAATGGG 1694
Db 71 AAAATTCATGGAAGCAATGGG 91
```

```

RESULT 4
BX933597          896 bp      mRNA      linear      VRT 02-FEB-2004
LOCUS             Gallus gallus finished cDNA, clone CHEST153f19.
DEFINITION
ACCESSION        BX933597
VERSION          BX933597.1  GI:41634125
KEYWORDS
SOURCE
ORGANISM
Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 896)
Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
Cranning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
Tickle,C. and Wilson,S.A.
Direct Submission
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chickest@hms.umist.ac.uk
BSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project.
This sequence is from the
BSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Bosch. cDNA was prepared
from RNA extracted from small intestine, and poly A-trimmed.
EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
coli DH10B.
FEATURES
source      Location/Qualifiers
1..896
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Compton Line 151"
    /db_xref="taxon:9031"
    /clone="CHEST153f19"
    /clone_lib="CSEOCHL18"
    /dev_stage="adult"
ORIGIN
Query Match      3.3%; Score 78; DB 5; Length 896;
Best Local Similarity 100.0%; Pred. No. 5.3e-28;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1617 AGACAGAAAGATGGCATTTACGGTACTTGGAAAATAGAGAAAATGAGAACTATGAAAA 1676
Db 11 AGACAGAAAGATGGCATTTACGGTACTTGGAAAATAGAGAAAATGAGAACTATGAAAA 70
QY 1677 ATTCATGGAAGCAATGGG 1694
Db 71 ATTCATGGAAGCAATGGG 88
RESULT 5
AC121429          198504 bp      DNA      linear      HTG 12-OCT-2002
LOCUS             Rattus norvegicus clone CH230-320M11, *** SEQUENCING IN PROGRESS
DEFINITION
ACCESSION        AC121429
VERSION          AC121429.3  GI:23907792
KEYWORDS          HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 198504)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooke,S., Amin,A., Anguitano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

```

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,M., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,D., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulaeged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,N., Muidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelimeh,O., Okuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkocch,C., Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quito,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajic,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,B., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weites,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission  
 Unpublished  
 2 (bases 1 to 198504)  
 Worley,K.C.  
 Direct Submission  
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 198504)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Oct 12, 2002 this sequence version replaced gi:21908516.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GXVD

Center clone name: CH230-320M11

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 176501 bases at least Q40

Consensus quality: 178427 bases at least Q30

Consensus quality: 179698 bases at least Q20

Estimated insert size: 183031; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 198504: contig of 198504 bp in length.

Location/Qualifiers

1. 198504

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-320M11"

1125..1981

/note="clone\_boundary"

clone\_end:Sp6

site:MboI

end\_sequence:RXANE78TV

101134..102773

/note="wgs contig"

192677..194505

/note="wgs contig"

complement(196385..197249)

/note="clone\_boundary"

clone\_end:T7

site:MboI

end\_sequence:RXANE78TJ

#### ORIGIN

Query Match 1.1%; Score 27; DB 2; Length 198504;

Best Local Similarity 100.0%; Pred. NO. 0.044;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2293 TTAATTGATTATTATTATTTT 2319

|||||

Db 117554 TTAATTGATTATTATTATTTT 117580

#### RESULT 6

AC107353

LOCUS

DEFINITION Rattus norvegicus clone CH230-47B12, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 2 unordered pieces.

AC107353

VERSION

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 240387)

AUTHORS

Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorusnewa, L., Louise, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okunolu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaje, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 240387)

Worley, K.C.

Direct Submission

Submitted (19-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240387)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 18, 2002 this sequence version replaced gi:21737464.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome



shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPEP
Center clone name: CH230-47B12
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 209782 bases at least Q40
Consensus quality: 213003 bases at least Q30
Consensus quality: 215205 bases at least Q20
Estimated insert size: 232384; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 2 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
*
* 1 171950: contig of 171950 bp in length
* 171951 172050: gap of unknown length
* 172051 240387: contig of 68337 bp in length.
FEATURES             source
    1..240387
       /organism="Rattus norvegicus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10116"
       /clones="CH230-47B12"
    misc_feature      1..1432
                       /note="wgs_contig"
    misc_feature      64734..66070
                       /note="wgs_contig"
ORIGIN
Query Match          1.1% Score 27; DB 2; Length 240387;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2293 TTAATTGATTATTATTATTTT 2319
Db 83111 TTAATTGATTATTATTATTTT 83137

RESULT 7
AC003695/c
LOCUS AC003695 115199 bp DNA linear PRI 29-OCT-1998
DEFINITION Homo sapiens chromosome 17, clone hRPC.859_O_20, complete sequence.
ACCESSION AC003695
VERSION AC003695.1 GI:3808090
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 115199)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPC.859_O_20
Unpublished
2 (bases 1 to 115199)
Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
Richardson,P., Lander,E., Allen,N., Baldwin,J., Barna,N.,
Beckerly,R., Cantu,C., Castle,A., Cooke,P., Daly,M.J., Depayre,E.,
Devon,K., Dewar,K., Dukette,B., Ferreira,P., Forrest,C., Gage,D.,
Gardyna,S., Gensheimer,S., Geraigery,K., Gilmartin,T., Gray,D.,

```

```

Hagos,B., Halphen,I., Harris,K., Horton,L., Howland,J.C., Huang,J.,
Hui,L., Jacotot,L., Linton,L., Mackenzie,J., Marquis,N.,
McDermott,J., McEwan,P., McGurk,A., Meldrim,J., Molla,M.,
Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A.,
Sarnaik,A., Shiu,P., Shyam,R., Stilwell,J., Stone,C., Strickland,C.,
Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wu,Y.,
Ye,W.J., Zentseva,I., Zhao,J. and Zody,M.
Direct Submission
Submitted (22-DEC-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 115199)
REFERENCE
AUTHORS
    Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
    Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
    Boutwell,C., Brown,A., Byrne,S., Cantu,C., Castle,A., Cerny,J.,
    Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Dukette,B.,
    Etemadi,S., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
    Gage,D., Gardyna,S., Gensheimer,S., Geraigery,K., Gilmartin,T.,
    Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L.,
    Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P.,
    McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
    Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,J., O'Connor,T.,
    Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A.,
    Shyam,R., Strange-Thomann,N., Stilwell,J., Stone,C., Strickland,C.,
    Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A.,
    Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (09-MAY-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 115199)
REFERENCE
AUTHORS
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
    Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
    Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
    Donegan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
    Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heatford,A.,
    Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
    Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
    McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
    Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
    O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
    Severy,P., Strange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
    Subramanian,A., Tesfaye,S., Tichovolovskiy,N., Torruella-Miller,I.,
    Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
    Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (29-OCT-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 29, 1998 this sequence version replaced gi:3126779.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             source
    1..115199
       /organism="Homo sapiens"
       /mol_type="genomic DNA"
       /db_xref="taxon:9606"
       /chromosome="17"
       /map="17"
       /clone="hRPC.859_O_20"
       /clone_lib="Peter de Jong/ human PAC library"
       complement(1..1236)
       /rpt_family="L1MC4"
       complement(1237..1543)
       /rpt_family="AlusX"
       complement(1544..1781)
       /rpt_family="L1MC4"
       2628..2654
       /rpt_family="GC rich"
       complement(3693..3969)
       /rpt_family="AlusX"
       4098..4197
       /rpt_family="MIR"

```

```

repeat_region 4842. .4898
/rpt_family="MIR"
complement(5392. .5702)
/rpt_family="AluSx"
5830. .6156
/rpt_family="L2"
complement(6636. .6816)
/rpt_family="MIR"
8340. .8486
/rpt_family="L2"
8483. .8713
/rpt_family="MIR"
complement(8730. .8803)
/rpt_family="MER69"
complement(8798. .8934)
/rpt_family="MER69"
9341. .9830
/rpt_family="AluSc"
complement(10554. .10596)
/rpt_family="(CA)n"
11190. .11469
/rpt_family="AluJo"
11484. .11516
/rpt_family="AT rich"
11798. .12100
/rpt_family="AluSc"
12209. .12285
/rpt_family="L1P"
complement(12339. .12524)
/rpt_family="AluSx"
12657. .12678
/rpt_family="AT rich"
12852. .13131
/rpt_family="L1PA8"
15019. .15092
/rpt_family="L1"
15389. .15447
/rpt_family="AT rich"
15574. .15691
/rpt_family="AT rich"
15762. .15982
/rpt_family="L1ME"
15983. .16280
/rpt_family="AluSg"
16281. .16558
/rpt_family="L1ME"
16576. .16877
/rpt_family="AluSc"
16918. .17350
/rpt_family="L1ME"
complement(18702. .19067)
/rpt_family="L1R37B"
complement(19077. .19369)
/rpt_family="AluSg"
19508. .19662
/rpt_family="MIR"
20325. .20380
/rpt_family="MIR"
complement(20484. .20598)
/rpt_family="MIR"
complement(21076. .21157)
/rpt_family="MLR1J"
21158. .21209
/rpt_family="(CA)n"
complement(21210. .21530)
/rpt_family="MLR1J"
complement(23483. .23703)
/rpt_family="L2"
complement(25009. .25079)
/rpt_family="MIR"
complement(25825. .25919)
/rpt_family="MIR"
26036. .26080

```

```

repeat_region /rpt_family="(CA)n"
complement(26566. .26672)
/rpt_family="MIR"
27224. .27535
/rpt_family="AluY"
27919. .27958
/rpt_family="AT rich"
complement(29206. .29526)
/rpt_family="MLR1J"
complement(29588. .29698)
/rpt_family="MIR"
30291. .30338
/rpt_family="(GGA)n"
complement(31067. .31124)
/rpt_family="(TAA)n"
complement(31125. .31412)
/rpt_family="AluSx"
complement(35239. .35741)
/rpt_family="L2"
complement(35742. .36039)
/rpt_family="AluSc"
complement(36040. .36867)
/rpt_family="L2"
complement(37547. .37641)
/rpt_family="(CA)n"
38996. .39047
/rpt_family="(CACG)n"
39324. .39366
/rpt_family="L2"
41170. .41422
/rpt_family="L1MD3"
41432. .41465
/rpt_family="(CA)n"
41469. .41672
/rpt_family="L1M2"

```

Query Match 1.1%; Score 26; DB 9; Length 115199;  
 Best Local Similarity 100.0%; Pred.No. 0.15;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1265 GCACAAAGAAAATGTCATACATTC 1290  
 |||||  
 Db 103587 GCACAAAGAAAATGTCATACATTC 103562

RESULT 8  
 AC009451/c  
 LOCUS 133455 bp DNA linear PRI 14-DEC-2002  
 DEFINITION Homo sapiens chromosome 17, clone CTD-2286H12, complete sequence.  
 ACCESSION AC009451  
 VERSION AC009451.21 GI:26801280  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 133455)  
 AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 17, clone CTD-2286H12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 133455)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.;  
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
 Castile,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,  
 Doneelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
 Funke,B., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
 Karas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
 Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
 Peterson,K., Pollara,V., Rile,R., Roberts,D., Roy,A., Severy,P.,

```

Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tsefaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE
JOURNAL
Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 133455)
REFERENCE
AUTHORS
Barren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tsefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Vieler,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE
JOURNAL
Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 133455)
REFERENCE
AUTHORS
Barren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tsefaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Vieler,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE
JOURNAL
Submitted (14-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 14, 2002 this sequence version replaced gi:23322752.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L892
Center clone name: 2286_H_12
-----

Only the first 133.5 kilobases of this clone are being submitted.
The remainder is overlapped by accession number AC011061 (WICGR
project L3268).
FEATURES
source
1..133455
    Location/Qualifiers
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="17"

```

```

repeat_region complement(15894..16094)
repeat_region /rpt_family="MIR"
repeat_region 16242..16384
repeat_region /rpt_family="MER5B"
repeat_region 16387..16439
repeat_region /rpt_family="L3"
repeat_region complement(18647..18788)
repeat_region /rpt_family="AluJb"
repeat_region complement(18793..19104)
repeat_region /rpt_family="AluSx"
repeat_region complement(19119..19255)
repeat_region /rpt_family="FLAM C"
repeat_region complement(20573..20872)
repeat_region /rpt_family="AluY"
unsure 21007..21035
repeat_region /note="<30 qual. SNGL region"
repeat_region 21196..21318
repeat_region /rpt_family="LTR16A"
repeat_region complement(21342..21623)
repeat_region /rpt_family="L2"
repeat_region 21865..21892
repeat_region /rpt_family="(CCCCCG)n"
repeat_region 24380..24665
repeat_region /rpt_family="AluSc"
repeat_region complement(25270..25603)
repeat_region /rpt_family="AluSc"
repeat_region complement(25673..25915)
repeat_region /rpt_family="MTIL"

Query Match 1.1%; Score 26; DB 9; Length 133455;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1265 GCACAAAGAAATGTCATACATTC 1290
Db 5703 GCACAAAGAAATGTCATACATTC 5678

RESULT 9
CR626936
LOCUS
DEFINITION Danio rerio clone DKEY-107122, WORKING DRAFT SEQUENCE, 5 unordered
pieces
ACCESSION CR626936
VERSION CR626936.3 GI:51090206
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 179668)
Burton, J.
Direct Submission
Submitted (08-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk
On Aug 9, 2004 this sequence version replaced gi:50896321.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zki07122
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 177410 bases at least Q40
Consensus quality: 177759 bases at least Q30
Consensus quality: 178052 bases at least Q20
Insert size: 179268; sum-of-contigs
Insert size: 173635; 3.1% error; agarose-fp

```

Quality coverage: 9.22x in Q20 bases; sum-of-contigs Quality coverage: 9.90x in Q20 bases; agarose-fp

-----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 5950: contig of 5950 bp in length  
 \* 5951 6050: gap of 100 bp  
 \* 6051 72517: contig of 66467 bp in length  
 \* 72518 72617: gap of 100 bp  
 \* 72618 87634: contig of 15017 bp in length  
 \* 87635 87735: gap of 100 bp  
 \* 87735 130094: contig of 42360 bp in length  
 \* 130095 130194: gap of 100 bp  
 \* 130195 179668: contig of 49474 bp in length.

## FEATURES

## source

1..179668  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-107122"  
 /clone\_lib="DankioKey"

## misc\_feature

1..5950  
 /note="assembly\_fragment:00021"

fragment\_chain:1

## misc\_feature

6051..72517  
 /note="assembly\_fragment:01524"

fragment\_chain:1

## misc\_feature

72618..87634  
 /note="assembly\_fragment:00072"

fragment\_chain:1

## misc\_feature

87735..130094  
 /note="assembly\_fragment:00264"

fragment\_chain:1

## misc\_feature

130195..179668  
 /note="assembly\_fragment:00871.0"

fragment\_chain:1

## ORIGIN

Query Match 1.1%; Score 26; DB 2; Length 179668;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2296 ATTCATTATTATTATTTTATTTTATTA 2321  
 Db 90229 ATTCATTATTATTATTTTATTTTATTA 90254

## RESULT 10

## AC025959

LOCUS AC025959 199534 bp DNA linear HTG 24-AUG-2002  
 DEFINITION Homo sapiens chromosome 17 clone RP11-350B20 map 17, WORKING DRAFT SEQUENCE, 34 unordered pieces.

## ACCESSION

AC025959

## VERSION

AC025959.3 GI:8072618

## KEYWORDS

HTG; HTGS PHASE1; HTGS\_DRAFT.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 199534)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lakoque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melgrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 199534)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavskiy, I., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lakoque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melgrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7407980.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L5511  
Center clone name: 350 B.20  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 181671 bases at least Q40  
Consensus quality: 190519 bases at least Q30  
Consensus quality: 193834 bases at least Q20  
Insert size: 172000; agarose-fp  
Insert size: 196234; sum-of-contigs  
Quality coverage: 4.3 in Q20 bases; agarose-fp  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1120: contig of 1120 bp in length  
1121 1220: gap of 100 bp  
1122 2554: contig of 1334 bp in length  
1221 2554: gap of 100 bp  
2555 3920: contig of 1266 bp in length  
3921 4020: gap of 100 bp  
4021 5363: contig of 1343 bp in length  
5364 6949: contig of 1486 bp in length  
6950 7049: gap of 100 bp  
7050 8867: contig of 1818 bp in length  
8868 8967: gap of 100 bp  
8968 11065: contig of 2098 bp in length  
11066 11492: gap of 100 bp  
11493 14592: contig of 3327 bp in length  
14593 15881: contig of 1289 bp in length  
15882 19837: contig of 3856 bp in length  
19838 22318: contig of 2381 bp in length  
22319 24419: gap of 100 bp  
24420 24541: contig of 2023 bp in length  
24542 28250: contig of 3709 bp in length  
28251 28350: gap of 100 bp  
28351 33280: contig of 4930 bp in length  
33281 33380: gap of 100 bp  
33381 37874: contig of 4494 bp in length  
37875 37974: gap of 100 bp  
37975 42401: contig of 4427 bp in length  
42402 46772: contig of 4271 bp in length  
46773 52797: contig of 5925 bp in length  
52798 52897: gap of 100 bp  
52898 57226: contig of 4329 bp in length  
57227 57326: gap of 100 bp  
57327 60434: contig of 3108 bp in length  
60435 60534: gap of 100 bp  
60535 66233: contig of 5699 bp in length  
66234 72227: contig of 5894 bp in length  
72228 72328: gap of 100 bp  
72329 80533: contig of 8106 bp in length  
80534 87266: contig of 6733 bp in length  
87267 87366: gap of 100 bp  
87367 92934: contig of 5568 bp in length  
92935 93034: gap of 100 bp  
93035 99776: contig of 6742 bp in length  
99777 99876: gap of 100 bp  
99877 107845: contig of 7969 bp in length  
107846 107945: gap of 100 bp  
107946 114972: contig of 7027 bp in length  
114973 115072: gap of 100 bp  
115073 123512: contig of 8440 bp in length  
123513 123612: gap of 100 bp  
123613 133090: contig of 9478 bp in length  
133091 133190: gap of 100 bp  
133191 143786: contig of 10596 bp in length  
143787 143886: gap of 100 bp  
143887 159481: contig of 15595 bp in length  
159482 159581: gap of 100 bp  
159582 177756: contig of 18175 bp in length  
177757 177856: gap of 100 bp  
177857 199534: contig of 21678 bp in length.

## FEATURES

source

Location/Qualifiers

1. .199534  
/organism="Homo sapiens"

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="17"
/map="17"
/clone_lib="RP11-RPCI-11 Human Male BAC"
1. .1120
/note="assembly_fragment"
1221. .2554
/note="assembly_fragment"
2655. .3920
/note="assembly_fragment"
4021. .5363
/note="assembly_fragment"
5464. .6949
/note="assembly_fragment"
7050. .8867
/note="assembly_fragment"
8968. .11065
/note="assembly_fragment"
11166. .14492
/note="assembly_fragment"
14593. .15881
/note="assembly_fragment
clone end:SP6
vector side:right"
15982. .19837
/note="assembly_fragment"
19938. .22318
/note="assembly_fragment"
22419. .24441
/note="assembly_fragment"
24542. .28250
/note="assembly_fragment"

Query Match
Best Local Similarity 1.1%; Score 26; DB 2; Length 199534;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1265 GCAACAAGAAAATGTCATACATTC 1290
|||||
Db 119106 GCAACAAGAAAATGTCATACATTC 119131

RESULT 11
AC025963/c
LOCUS AC025963 205195 bp DNA linear HTG 03-APR-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-798A22 map 17, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
ACCESSION AC025963
VERSION AC025963.2 GI:7387349
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205195)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

```

Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 3, 2000 this sequence version replaced GI:7262751.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5699  
 Center clone name: 798\_A22  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 181877 bases at least Q40  
 Consensus quality: 193636 bases at least Q30  
 Consensus quality: 198651 bases at least Q20  
 Insert size: 194000; agarose-fp  
 Insert size: 202495; sum-of-contigs  
 Quality coverage: 4.2 in Q20 bases; agarose-fp  
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 28 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1102: contig of 1102 bp in length  
 \* 1103 1202: gap of 100 bp  
 \* 1203 2273: contig of 1071 bp in length  
 \* 2274 2373: gap of 100 bp  
 \* 2374 3696: contig of 1323 bp in length  
 \* 3697 3796: gap of 100 bp  
 \* 3797 5006: contig of 1210 bp in length  
 \* 5007 5106: gap of 100 bp  
 \* 5107 6548: contig of 1442 bp in length  
 \* 6549 7939: contig of 1291 bp in length  
 \* 7940 8039: gap of 100 bp  
 \* 8040 9873: contig of 1834 bp in length  
 \* 9874 9973: gap of 100 bp  
 \* 9974 11841: contig of 1868 bp in length  
 \* 11842 11941: gap of 100 bp  
 \* 11942 13276: contig of 1335 bp in length  
 \* 13277 13376: gap of 100 bp  
 \* 13377 17279: contig of 3903 bp in length  
 \* 17280 21787: gap of 4408 bp in length  
 \* 21788 21887: gap of 100 bp  
 \* 21888 27031: contig of 5144 bp in length  
 \* 27032 27131: gap of 100 bp  
 \* 27132 30904: contig of 3773 bp in length  
 \* 30905 31004: gap of 100 bp  
 \* 31005 35764: contig of 4760 bp in length  
 \* 35765 42646: gap of 100 bp  
 \* 42646: contig of 6782 bp in length

TITLE  
 JOURNAL  
 COMMENT

```
* 42647 42746: gap of 100 bp
* 42747 49542: contig of 6796 bp in length
* 49543 49642: gap of 100 bp
* 49643 57524: contig of 7882 bp in length
* 57525 57624: gap of 100 bp
* 57625 65221: contig of 7597 bp in length
* 65222 65322: gap of 100 bp
* 65323 75663: contig of 10242 bp in length
* 75664 75664: gap of 100 bp
* 75665 86184: contig of 10521 bp in length
* 86185 86284: gap of 100 bp
* 86285 96173: contig of 9889 bp in length
* 96174 96273: gap of 100 bp
* 96274 108130: contig of 11857 bp in length
* 108131 108230: gap of 100 bp
* 108231 119983: contig of 11753 bp in length
* 119984 120083: gap of 100 bp
* 120084 132844: contig of 12761 bp in length
* 132845 132944: gap of 100 bp
* 132945 148624: contig of 15680 bp in length
* 148625 148724: gap of 100 bp
* 148725 164504: contig of 15780 bp in length
* 164505 181271: contig of 16667 bp in length
* 181272 181371: gap of 100 bp
* 181372 205195: contig of 23824 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..205195
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="17"
                /map="17"
                /clone="RP11-798A22"
                /clone_lib="RPC1-11 Human Male BAC"
            1..1102
                /note="assembly_fragment"
            1203..2273
                /note="assembly_fragment"
            2374..3696
                /note="assembly_fragment"
            3797..5006
                /note="assembly_fragment"
            5107..6548
                /note="assembly_fragment"
            6649..7939
                /note="assembly_fragment"
            8040..9873
                /note="assembly_fragment"
            9974..11841
                /note="assembly_fragment"
            11942..13276
                /note="assembly_fragment
                clone_end:SP6
                vector_side:right"
            13377..17279
                /note="assembly_fragment"
            17380..21787
                /note="assembly_fragment"
            21888..27031
                /note="assembly_fragment"
            27132..30904
                /note="assembly_fragment"
            31005..35764
                /note="assembly_fragment"
            35865..42646
                /note="assembly_fragment"
            42747..49542
                /note="assembly_fragment"
            49643..57524
                /note="assembly_fragment"
            57625..65221
                /note="assembly_fragment"
```

```
misc_feature 65322..75563
              /note="assembly_fragment"
misc_feature 75664..86184
              /note="assembly_fragment"
misc_feature 86285..96173
              /note="assembly_fragment"
misc_feature 96274..108130
              /note="assembly_fragment"
misc_feature 108231..119983
              /note="assembly_fragment
              clone_end:17
              vector_side:left"
misc_feature 120084..132844
              /note="assembly_fragment"
misc_feature 132945..148624
              /note="assembly_fragment"
misc_feature 148725..164504
              /note="assembly_fragment"
misc_feature 164605..181271
              /note="assembly_fragment"
misc_feature 181372..205195
              /note="assembly_fragment"
```

## ORIGIN

```
Query Match      1.1%; Score 26; DB 2; Length 205195;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1265 GCACCAAGAAATGTCAATACATTC 1290
      |||||
DB 26137 GCACCAAGAAATGTCAATACATTC 26112
```

## RESULT 12

HS349A12/c

## LOCUS

DEFINITION

HS349A12 132948 bp DNA linear PRI 06-JUL-2000  
Human DNA sequence from clone RP3-349A12 on chromosome 6p21.31-22.2  
Contains part of a novel gene similar to KIAA0701 protein, the  
TAF2I (TATA box binding protein (TBP)-associated factor, RNA  
polymerase II, I, 28KD) gene, ESTs, STSs, GSSs and a CpG island,  
complete sequence.

## ACCESSION

AL033520.16

## VERSION

GI:8218054

## KEYWORDS

HTG; CpG island; KIAA0701; RNA polymerase; TAF2I; TATA box binding

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 (bases 1 to 132948)

## AUTHORS

Williams, S.

## TITLE

Direct Submission

## JOURNAL

Submitted (05-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jun 3, 2000 this sequence version replaced gi:7799905.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

RP3-349A12 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-349A12 The true right end of clone RP3-375P9 is at 5971 in this sequence.

## FEATURES

## source

```
Location/Qualifiers
1..132948
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="6"
/map="p21.31-22.2"
/clone="RP3-349A12"
/clone_lib="RPCI-3"
1..345
/note="match: STS: Em:G22496 Em:R11511"
211..512
/note="AluJo repeat: matches 1..302 of consensus"
688..884
/note="L1M3 repeat: matches 6107..6304 of consensus"
882..1200
/note="L1M1 repeat: matches 5422..5746 of consensus"
1201..1510
/note="AluY repeat: matches 1..308 of consensus"
1514..1803
/note="AluSx repeat: matches 1..291 of consensus"
1804..2654
/note="L1M1 repeat: matches 4601..5422 of consensus"
2656..2819
/note="AluJb repeat: matches 129..292 of consensus"
2838..3125
/note="AluY repeat: matches 1..287 of consensus"
3179..3305
/note="L1M1 repeat: matches 4477..4613 of consensus"
3307..3361
/note="11 copies 5 mer ccttc 92% conserved"
3384..3660
/note="AluSx repeat: matches 14..290 of consensus"
3663..5140
/note="L1M1 repeat: matches 3030..4494 of consensus"
5142..5436
/note="AluJb repeat: matches 14..310 of consensus"
5437..5523
/note="AluJb repeat: matches 205..293 of consensus"
5553..5849
/note="AluSx repeat: matches 1..297 of consensus"
5853..6215
/note="L1M1 repeat: matches 2351..2715 of consensus"
6216..6521
/note="AluSx repeat: matches 1..312 of consensus"
6523..6551
/note="AluJb repeat: matches 1..29 of consensus"
6552..7090
/note="L1M1 repeat: matches 1810..2351 of consensus"
7091..7387
/note="AluSp repeat: matches 7..298 of consensus"
7388..7934
/note="L1M1 repeat: matches 1261..1810 of consensus"
7935..8223
/note="AluJo repeat: matches 1..289 of consensus"
8224..10403
/note="L1M1 repeat: matches -1305..1261 of consensus"
10404..10693
/note="AluJb repeat: matches 1..302 of consensus"
10694..10779
/note="L1M1 repeat: matches -1390..-1305 of consensus"
10785..11090
/note="AluSx repeat: matches 1..307 of consensus"
```

## repeat\_region

## repeat\_region

## gene

## mRNA

## CDS

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

```
11164..11472
/note="AluSg repeat: matches 1..300 of consensus"
12037..12095
/note="L2 repeat: matches 2678..2733 of consensus"
16124..29943
/gene="dJ349A12.1"
join(<16124..16363,21313..21427,24454..24608,24695..24845,
28083..28345,28813..28921,29048..29154,29542..29943)
/gene="dJ349A12.1"
/product="dJ349A12.1 (similar to KIAA0701 protein)"
/note="match: ESTs: Em:AA705888"
/evidence=not_experimental
join(<16124..16363,21313..21427,24454..24608,24695..24845,
28083..28345,28813..28921,29048..29154,29542..29667)
/gene="dJ349A12.1"
/note="match: proteins: Tr:O75183"
/codon_start=1
/evidence=not_experimental
/product="dJ349A12.1 (similar to KIAA0701 protein)"
/protein_id="CAB96724.1"
/db_xref="GI:8953444"
/db_xref="Tr:EMBL:Q9NQF5"
/translation="RLQQLTKDTSMTGSPLOQTACIGVLPFAEVALMHHPAGPA
VDASAGSDSLVDSLSUSDRELDKSDASDQPTVSQSFDSGLDSSGPDRI
SVDSGDSFVNLJESGSPSPFGLSVNSVNDAGVQGPLVNNYGGSPAAANSVS
PSGDLIFHPVSLVKVNEYSFGIEVRGDLTVLQAEELTLOQLGTGVPAAHLHGQ
CPGTCFQBSSTLKTGHIRPAVLGFVGPAAVHSPASONGFLHLLHGGCDLLTTS
VLSGLGPFLEDEEIPVIVPMOIELNLSITLKDDIPPIYTPSPGPIITLAMEHVLR
RSDGVFHGAADKPSAEVLKSEKQPPKEQVFLVPTGEVFEQVKELPILQKELI
ETKQLANANQDKELQELIRKYNPFPEL"
17011..17135
/note="FLAM A repeat: matches 1..127 of consensus"
17237..17375
/note="MIR repeat: matches 52..189 of consensus"
17430..17525
/note="L2 repeat: matches 2617..2709 of consensus"
17785..18001
/note="MIR repeat: matches 47..262 of consensus"
18104..18217
/note="L2 repeat: matches 1323..1435 of consensus"
18367..18654
/note="AluJb repeat: matches 1..299 of consensus"
18975..19113
/note="AluJb repeat: matches 1..134 of consensus"
19326..19632
/note="AluY repeat: matches 1..309 of consensus"
19748..19781
/note="MER5A repeat: matches 132..163 of consensus"
19782..20083
/note="AluY repeat: matches 1..302 of consensus"
20084..20141
/note="MER5A repeat: matches 64..132 of consensus"
20142..20435
/note="AluJb repeat: matches 1..296 of consensus"
20476..20499
/note="MIR repeat: matches 121..144 of consensus"
20789..20893
/note="L2 repeat: matches 2614..2747 of consensus"
21915..22209
/note="AluSx repeat: matches 1..297 of consensus"
22225..22991
/note="L1M7 repeat: matches 5423..6287 of consensus"
23039..23156
/note="L1M5 repeat: matches 7824..7935 of consensus"
23159..23228
/note="AluS repeat: matches 2..71 of consensus"
23362..23656
/note="AluS repeat: matches 1..304 of consensus"
23658..23791
/note="L1M5 repeat: matches 7716..7830 of consensus"
23792..24231
/note="L1M3 repeat: matches 5733..6184 of consensus"
24232..24339
```



```

/note="L1LMCS repeat: matches 7603. .7716 of consensus"
25255. .25541
/note="AluSq repeat: matches 1. .289 of consensus"
25566. .25669
/note="AluJo repeat: matches 1. .93 of consensus"
25670. .25970
/note="AluSx repeat: matches 1. .300 of consensus"
25971. .26127
/note="AluJo repeat: matches 93. .288 of consensus"
26224. .26530
/note="AluSx repeat: matches 1. .309 of consensus"
26689. .26869
/note="L1ME1 repeat: matches 5775. .5962 of consensus"
26870. .27176
/note="AluSx repeat: matches 2. .310 of consensus"
27195. .27330

Query Match          1.0%; Score 25; DB 9; Length 132948;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2296 ATTGATTATTATTATTTT 2320
Db      5537 ATTGATTATTATTATTTT 5513

RESULT 13
AC141717
LOCUS      AC141717      141876 bp      DNA      linear      HTG 19-MAR-2003
DEFINITION Apis mellifera clone CH224-57G21, WORKING DRAFT SEQUENCE, 6
unordered pieces.
ACCESSION      AC141717.1 GI:291233901
VERSION      HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE      Apis mellifera (honey bee)
ORGANISM      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apoidea; Aculeata; Apoidea;
Apis;
1 (bases 1 to 141876)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Cartoll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karissone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Maesey,E., Mahoney,E., Mcleod,M.P., Meador,M., Mei,G., Mezker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogwu,M., Okwuonu,G.,
Oraquyne,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,

```

```

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umami,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 141876)
Worley,K.C.
Direct Submission
Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: AMFM
Center clone name: CH224-57G21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 140842 bases at least Q40
Consensus quality: 141533 bases at least Q30
Consensus quality: 142031 bases at least Q20
Estimated insert size: 140890; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3254: contig of 3254 bp in length
* 3255 3354: gap of unknown length
* 7237: contig of 3883 bp in length
* 7338 16684: contig of 9347 bp in length
* 16685 16784: gap of unknown length
* 16785 26794: contig of 10010 bp in length
* 26795 26894: gap of unknown length
* 26895 66946: contig of 40052 bp in length
* 66947 67046: gap of unknown length
* 67047 141876: contig of 74830 bp in length.
*
* Location/Qualifiers
* 1.141876
* /organism="Apis mellifera"
* /mol_type="genomic DNA"
* /db_xref="taxon:7460"
* /clone="CH224-57G21"
*
FEATURES
source
ORIGIN
Query Match          1.0%; Score 25; DB 2; Length 141876;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2296 ATTGATTATTATTATTTT 2320
Db      2299 ATTGATTATTATTATTTT 2323

RESULT 14
AC068973

```

```

LOCUS      AC068973      166622 bp      DNA      linear      HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-241P12, WORKING DRAFT
ACCESSION AC068973
VERSION   HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS  Homo sapiens (human)
SOURCE    Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166622)
           Waterston,R.H.
           The sequence of Homo sapiens clone
           Unpublished
REFERENCE 2 (bases 1 to 166622)
           Waterston,R.H.
           Direct Submission
           Submitted (14-MAY-2000) Genome Sequencing Center, Washington
           University School of Medicine, 444 Forest Park Parkway, St. Louis,
           MO 63108, USA
           On Jun 12, 2000 this sequence version replaced gi:7801482.
COMMENT   ----- Genome Center -----
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc/index.shtml
           ----- Project Information -----
           Center project name: H.NH0241P12
           ----- Summary Statistics -----
           Sequencing vector: M13; 100%
           Chemistry: Dye-terminator; 100% of reads
           Assembly: Dye-terminator Big Dye; 0% of reads
           Consensus quality: 147920 bases at least Q40
           Consensus quality: 154306 bases at least Q30
           Consensus quality: 157356 bases at least Q20
           Insert size: 154000; agarose-fp
           Insert size: 163922; sum-of-contigs
           Quality coverage: 3.38 in Q20 bases; agarose-fp
           Quality coverage: 3.25 in Q20 bases; sum-of-contigs
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 28 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence.
           * as soon as it is available and the accession number will
           * be preserved.
           -----
           1 1692: contig of 1692 bp in length
           * 1693
           * 1793: gap of unknown length
           * 3522: contig of 1730 bp in length
           * 3523: gap of unknown length
           * 3623: contig of 1955 bp in length
           * 5578: gap of unknown length
           * 5678: contig of 1613 bp in length
           * 7291: gap of unknown length
           * 7391
           * 10182: contig of 2792 bp in length
           * 10183
           * 10283: gap of unknown length
           * 10283
           * 14325: contig of 4043 bp in length
           * 14326
           * 14426: gap of unknown length
           * 16456: contig of 2031 bp in length
           * 16557: gap of unknown length
           * 16557
           * 19681: contig of 3125 bp in length
           * 19782
           * 23298: contig of 3517 bp in length
           * 23299
           * 27567: gap of unknown length
           * 27568
           * 30974: contig of 4169 bp in length
           * 30975
           * 31074: gap of unknown length
           * 31075
           * 31075: contig of 3064 bp in length
           *
           34238: gap of unknown length
           38278: contig of 4040 bp in length
           38378: gap of unknown length
           41146: contig of 2768 bp in length
           41246: gap of unknown length
           46056: contig of 4810 bp in length
           46156: gap of unknown length
           52614: contig of 6458 bp in length
           52714: gap of unknown length
           58587: contig of 5873 bp in length
           58588
           65097: contig of 6410 bp in length
           65198
           73880: contig of 8683 bp in length
           73980: gap of unknown length
           80729: contig of 6749 bp in length
           80829: gap of unknown length
           86906: contig of 6077 bp in length
           86907
           95091: contig of 8085 bp in length
           95191: gap of unknown length
           104083: contig of 8892 bp in length
           104084
           115396: contig of 11213 bp in length
           115397
           124073: contig of 8577 bp in length
           124174
           135137: contig of 10964 bp in length
           135138
           135237: gap of unknown length
           145807: contig of 10570 bp in length
           145808
           166622: contig of 20715 bp in length.
           Location/Qualifiers
             1..166622
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="8"
               /clone="RP11-241P12"
             1..1692
               /note="assembly_name:Contig3"
             1793..3522
               /note="assembly_name:Contig5"
             3623..5577
               /note="assembly_name:Contig6"
             5678..7290
               /note="assembly_name:Contig7"
             7391..10182
               /note="assembly_name:Contig8"
             10283..14325
               /note="assembly_name:Contig9"
             14426..16456
               /note="assembly_name:Contig10"
             16557..19681
               /note="assembly_name:Contig11"
             19782..23298
               /note="assembly_name:Contig12"
             23399..27567
               /note="assembly_name:Contig13"
             27668..30974
               /note="assembly_name:Contig14"
             31075..34138
               /note="assembly_name:Contig15"
             clone_end:SP6
             vector_side:left
             34239..38278
               /note="assembly_name:Contig16"
             38379..41146
               /note="assembly_name:Contig17"
             41247..46056
               /note="assembly_name:Contig18"
             46157..52614
               /note="assembly_name:Contig19"
           -----
           FEATURES
           source
             misc_feature
               1793..3522
                 /note="assembly_name:Contig3"
             misc_feature
               3623..5577
                 /note="assembly_name:Contig5"
             misc_feature
               5678..7290
                 /note="assembly_name:Contig6"
             misc_feature
               7391..10182
                 /note="assembly_name:Contig7"
             misc_feature
               10283..14325
                 /note="assembly_name:Contig8"
             misc_feature
               14426..16456
                 /note="assembly_name:Contig9"
             misc_feature
               16557..19681
                 /note="assembly_name:Contig10"
             misc_feature
               19782..23298
                 /note="assembly_name:Contig11"
             misc_feature
               23399..27567
                 /note="assembly_name:Contig12"
             misc_feature
               27668..30974
                 /note="assembly_name:Contig13"
             misc_feature
               31075..34138
                 /note="assembly_name:Contig14"
             misc_feature
               clone_end:SP6
             misc_feature
               34239..38278
                 /note="assembly_name:Contig16"
             misc_feature
               38379..41146
                 /note="assembly_name:Contig17"
             misc_feature
               41247..46056
                 /note="assembly_name:Contig18"
             misc_feature
               46157..52614
                 /note="assembly_name:Contig19"
           -----

```



```
repeat_region /rpt_family="MLT1J2"
repeat_region complement(1592..1655)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(1675..1844)
repeat_region /rpt_family="L1M4"
repeat_region complement(1860..2187)
repeat_region /rpt_family="MER76"
repeat_region complement(2190..2598)
repeat_region /rpt_family="L1M4"
repeat_region 4090..4321
repeat_region /rpt_family="MER1B"
repeat_region complement(4751..5037)
repeat_region /rpt_family="AluJo"
repeat_region 6251..6291
repeat_region /rpt_family="GC_rich"
repeat_region 7341..7504
repeat_region /rpt_family="MER5A"
repeat_region 8072..8096
repeat_region /rpt_family="(A)n"
repeat_region 9408..9791
repeat_region /rpt_family="THE1C"
repeat_region 10075..10393
repeat_region /rpt_family="L1MC4a"
repeat_region 10963..11107
repeat_region /rpt_family="L1MC4a"
repeat_region complement(11124..11360)
repeat_region /rpt_family="MIR"
repeat_region 11505..12283
repeat_region /rpt_family="L1MC4a"
repeat_region 15377..15741
repeat_region /rpt_family="MLT1A1"
repeat_region 17846..17906
repeat_region /rpt_family="(TATAA)n"
repeat_region complement(17909..18160)
repeat_region /rpt_family="AluSx"
repeat_region complement(18412..18708)
repeat_region /rpt_family="AluSq"
repeat_region complement(18755..19749)
repeat_region /rpt_family="L1MC2"
repeat_region complement(19750..20046)
repeat_region /rpt_family="AluSx"
repeat_region complement(20047..21283)
repeat_region /rpt_family="L1MC2"
repeat_region complement(21284..21671)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(21672..21724)
repeat_region /rpt_family="L1MC2"
repeat_region complement(21731..21785)
repeat_region /rpt_family="L1P4"
repeat_region complement(22284..23572)
repeat_region /rpt_family="L1MD1"
repeat_region complement(23627..23725)
repeat_region /rpt_family="MER70B"
repeat_region complement(24178..25190)
repeat_region /rpt_family="L1MD1"
repeat_region 25200..25374
repeat_region /rpt_family="MER58A"
repeat_region complement(25931..26241)
repeat_region /rpt_family="AluY"
repeat_region complement(26510..26818)
repeat_region /rpt_family="AluSx"
repeat_region 27884..28064
repeat_region /rpt_family="MIR"
repeat_region 28711..28925
repeat_region /rpt_family="L1P10"
repeat_region complement(29342..29474)
repeat_region /rpt_family="FLAM_C"
repeat_region 30069..30126
repeat_region /rpt_family="CT-rich"
repeat_region complement(31024..31316)
repeat_region /rpt_family="AluJb"
repeat_region 31677..31770
repeat_region /rpt_family="MIR"
```

```
repeat_region 32166..32193
repeat_region /rpt_family="AT_rich"
repeat_region complement(32384..32699)
repeat_region /rpt_family="AluJo"
repeat_region complement(33353..33657)
repeat_region /rpt_family="AluSx"
repeat_region complement(33673..33744)
repeat_region /rpt_family="tRNA-Cys-TGC"
repeat_region 33793..34079
repeat_region /rpt_family="AluJo"
repeat_region 34649..34761
repeat_region /rpt_family="L2"
repeat_region complement(34732..34850)
repeat_region /rpt_family="L2"
repeat_region 36383..36627
repeat_region /rpt_family="MIR"
repeat_region 36771..36797
repeat_region /rpt_family="AT_rich"
```

Query Match 1.0%; Score 25; DB 9; Length 167118;  
Best Local Similarity 100.0%; Pred. No. 0.47;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2298 TCATTATTATTATTATTTTTTTTAG 2322  
|||  
Db 991 TGATTATTATTATTTTTTTTAG 967

Search completed: January 14, 2005, 22:40:28  
Job time : 10076 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 14, 2005, 12:21:52 ; Search time 1130 Seconds  
(without alignments)  
11060.948 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agctctctgcgcagaaagg.....atctgaagctcactatttcag 2381

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 4134886 seqs, 2624710521 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2381	100.0	2381	11	ADL90127 Chicken i
2	336	14.1	336	11	ADL90128 Chicken i
3	24	1.0	1160	3	AAF22348 Human sec
4	24	1.0	1160	10	ADC20223 Human sec
5	24	1.0	1160	10	ABT16829 Human sec
6	24	1.0	1160	10	ABZ67143 Human sec
7	24	1.0	6130	4	AAS46723 Tumour su
8	24	1.0	6130	6	AAS61411 Human gen
9	24	1.0	7133	4	AAS46388 Tumour su
C 10	24	1.0	22901	4	ABL19488 Drosophil
C 11	24	1.0	226475	9	AAD58279 Human tum
C 12	23	1.0	23	11	ADL90141 Chicken i
13	23	1.0	472	8	ABX39142 Bovine ES
14	23	1.0	503	6	ABV96362 Human pan
15	23	1.0	1112	10	ADC20846 Human sec
16	23	1.0	1112	10	ABT16983 Human sec
17	23	1.0	1112	10	ABZ67931 Human sec
18	23	1.0	1238	10	ADC20845 Human sec
19	23	1.0	1238	10	ABT16982 Human sec
20	23	1.0	1238	10	ABZ67930 Human sec
C 21	23	1.0	5054	4	AAK83477 Human imm

22	23	1.0	6621	6	ABL32912	Human imm
C 23	23	1.0	8696	4	ABL05912	Drosophil
C 24	23	1.0	10139	5	AAS29226	Genomic s
C 25	23	1.0	10139	6	ABS68366	Human DNA
C 26	23	1.0	10139	10	ADC25488	Human cDN
C 27	23	1.0	133100	12	ADP45594	Human NUM
C 28	23	1.0	175737	6	ABK83571	Human cDN
C 29	23	1.0	175737	10	ADL13596	Osteoearth
C 30	23	1.0	175737	12	ADQ18934	Human sof
C 31	22	0.9	22	11	ADL90139	Chicken i
C 32	22	0.9	22	11	ADL90137	Chicken i
C 33	22	0.9	22	11	ADL90143	Chicken i
C 34	22	0.9	533	10	ADD26966	Human adi
C 35	22	0.9	1299	2	AAQ99306	Genomic 5
C 36	22	0.9	2000	6	ABZ16608	Arabidops
C 37	22	0.9	3852	12	ADP74637	Nucleotid
38	22	0.9	3975	1	AAAN81157	Malaria-s
39	22	0.9	3975	2	AAQ22999	SERP gene
40	22	0.9	4274	6	ABL34066	Human imm
41	22	0.9	5388	6	ABK31344	Signal tr
42	22	0.9	5388	6	ABL70305	Chemical
43	22	0.9	5388	6	AAS61239	Human gen
C 44	22	0.9	6012	6	ABN79994	Human che
45	22	0.9	6076	4	AAS46663	Tumour su

ALIGNMENTS

RESULT 1  
ADL90127  
ID ADL90127 standard; DNA; 2381 BP.  
XX AC ADL90127;  
XX DT 20-MAY-2004 (first entry)  
XX DE Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region.  
XX KW Chicken; ds; intestinal fatty acid binding protein; iFABP;  
XX KW gut specific promoter; transgenic.  
XX OS Gallus gallus.  
XX PN US2003177516-A1.  
XX PD 18-SEP-2003.  
XX PF 14-MAR-2002; 2002US-00099663.  
XX PR 14-MAR-2002; 2002US-00099663.  
XX PA (HORS/) HORSEMAN N D.  
XX PRAT(//) PRATT S L.  
XX PI Horseman ND, Pratt SL;  
XX DR WPI; 2003-898653/82.  
XX PT New nucleic acid molecule comprising an isolated avian gut-specific gene  
XX PT expression control region, useful for regulating heterologous nucleic  
XX PT acids in transgenic avians, and for generating transgenic birds.  
XX PS Claim 1; SEQ ID NO 1; 28pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising an isolated  
XX CC avian gut-specific gene expression control region appearing as  
XX CC ADL90127(Chicken intestinal fatty acid binding protein, iFABP, gene, 5'  
XX CC region or ADL90128 (Chicken iFABP promoter) or its degenerate variant.  
XX CC Also included are a recombinant DNA molecule comprising an isolate avian  
XX CC gut-specific gene expression control region operably linked to a nucleic  
XX CC acid insert encoding a polypeptide, an expression vector that integrates  
XX CC into a host cell (and comprising the isolated avian gut-specific gene

CC expression control region), expressing a heterologous polypeptide in a  
 CC host cell (by transfecting a eukaryotic cell with the recombinant DNA  
 CC molecule, and culturing the transfected cell in a medium suitable for  
 CC expression of a heterologous polypeptide under the control of an avian  
 CC intestinal fatty acid binding protein (iFABP) or cp35 gene expression  
 CC control region encoded by the recombinant DNA molecule), a eukaryotic  
 CC cell transformed with the expression vector (or its progeny, which  
 CC expresses a heterologous polypeptide) and a transgenic avian having a  
 CC heterologous polynucleotide sequence comprising the nucleic acid insert.  
 CC The nucleic acids are useful for regulating heterologous nucleic acids in  
 CC transgenic avians, as probes in nucleic acid hybridisation assays for  
 CC detecting the iFABP gene expression control region, and for generating  
 CC transgenic birds. The present sequence is the Chicken intestinal fatty  
 CC acid binding protein, iFABP, gene, 5' region.  
 XX  
 SQ Sequence 2381 BP; 775 A; 428 C; 469 G; 709 T; 0 U; 0 Other;

Query Watch		100.0%;	Score 2381;	DB 11;	Length 2381;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2381;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGCTTCCTCGCGCAGAAAAGGCTGTGGGTTCTTGTTCCTCACACAGCTTAAGCAAAATC	60		
DB	1	AGCTTCCTCGCGCAGAAAAGGCTGTGGGTTCTTGTTCCTCACACAGCTTAAGCAAAATC	60		
QY	61	CCCAAGTTCCAAACGTCGGCTGTAAAGAGAGATGGCTCCTTCAAAATGAAGTGAATATG	120		
DB	61	CCCAAGTTCCAAACGTCGGCTGTAAAGAGAGATGGCTCCTTCAAAATGAAGTGAATATG	120		
QY	121	AAATTAATCATAAACGAGCTCTGTGGCAGATCAGAGATAACCTCTCTCGGCAAAAT	180		
DB	121	AAATTAATCATAAACGAGCTCTGTGGCAGATCAGAGATAACCTCTCTCGGCAAAAT	180		
QY	181	CTTAAAGTGATAGTAGACAGAGGCTGTGGCACTAACTTAATGCAATGAATGTAACACA	240		
DB	181	CTTAAAGTGATAGTAGACAGAGGCTGTGGCACTAACTTAATGCAATGAATGTAACACA	240		
QY	241	TTTGATCTTCTAGGAGCAAAAGGCTCTGGAAAAAATAATTTCTGGTCCACAGTCAGT	300		
DB	241	TTTGATCTTCTAGGAGCAAAAGGCTCTGGAAAAAATAATTTCTGGTCCACAGTCAGT	300		
QY	301	AGCAGCCTGTTTGGGTCGAACACTACAGCAACTTGTGGTGAACAAATCTAAGTTG	360		
DB	301	AGCAGCCTGTTTGGGTCGAACACTACAGCAACTTGTGGTGAACAAATCTAAGTTG	360		
QY	361	TTTTCCTTCTCTTCTTAACTTCTGTACAGTCTAAAGGTGAAGTAGTCTATTGAGT	420		
DB	361	TTTTCCTTCTCTTCTTAACTTCTGTACAGTCTAAAGGTGAAGTAGTCTATTGAGT	420		
QY	421	TACTTCCCTCTGCATCTCTTAGCCAGATTAGCAATTTGATTTCAAAATGAACCTGAGTGA	480		
DB	421	TACTTCCCTCTGCATCTCTTAGCCAGATTAGCAATTTGATTTCAAAATGAACCTGAGTGA	480		
QY	481	ATGGAAGCCACACTATTTTGGTATACACAGCAAGTCTTAATTTATAGTTATACCTCA	540		
DB	481	ATGGAAGCCACACTATTTTGGTATACACAGCAAGTCTTAATTTATAGTTATACCTCA	540		
QY	541	GTAAGACCTTTTCTGCGAGTCTGGAAGAAAAGAGATTATGATAACACAGACTAGTA	600		
DB	541	GTAAGACCTTTTCTGCGAGTCTGGAAGAAAAGAGATTATGATAACACAGACTAGTA	600		
QY	601	AAATTCATTAGTTAGAGCAACCTGTTATCTGTGTGATTAAGCAATTCATTTCAGCAT	660		
DB	601	AAATTCATTAGTTAGAGCAACCTGTTATCTGTGTGATTAAGCAATTCATTTCAGCAT	660		
QY	661	TCAGGATTTACATTTTGAAGCTAATAGACAGAGATTGGTGGCTCCATAGGAACAGA	720		
DB	661	TCAGGATTTACATTTTGAAGCTAATAGACAGAGATTGGTGGCTCCATAGGAACAGA	720		
QY	721	CTAAGCTAATCTGAGTTAGTACAGCAGATTTAGCACCAAGCAAAATTTGCTCAGTTTC	780		
DB	721	CTAAGCTAATCTGAGTTAGTACAGCAGATTTAGCACCAAGCAAAATTTGCTCAGTTTC	780		

QY	781	AAGTAGCAGTATCTTGTGGGGAAGAAGGAGCTGAGCCAGTGTGTGCTCATTTTCTGCATT	840
DB	781	AAGTAGCAGTATCTTGTGGGGAAGAAGGAGCTGAGCCAGTGTGTGCTCATTTTCTGCATT	840
QY	841	ATCCTTCAACATTTAAACCTGGGATCTATGGAAATCAAAACAGCTTGGGTAAATTCAT	900
DB	841	ATCCTTCAACATTTAAACCTGGGATCTATGGAAATCAAAACAGCTTGGGTAAATTCAT	900
QY	901	TAGCAGCAGATCACTACTGTAGGAATGGACAGAAACAGAGCATTCACTGAATGGGCTAT	960
DB	901	TAGCAGCAGATCACTACTGTAGGAATGGACAGAAACAGAGCATTCACTGAATGGGCTAT	960
QY	961	AATATAGAGAAATACGTAGAAAGGTCTCGAATTTAGACTACCTATTAAAGAGTGAAGACA	1020
DB	961	AATATAGAGAAATACGTAGAAAGGTCTCGAATTTAGACTACCTATTAAAGAGTGAAGACA	1020
QY	1021	CGAATGGAGAAATATCATCGCAATTTCTGTAGCTCAGCAGCTAGACTCGAAGGTTTCTGAAA	1080
DB	1021	CGAATGGAGAAATATCATCGCAATTTCTGTAGCTCAGCAGCTAGACTCGAAGGTTTCTGAAA	1080
QY	1081	CTGAACCGAGTTTCCCAAACTACCTGTGGATGTTTCTGTGGATGTTTCTCATGCTTAT	1140
DB	1081	CTGAACCGAGTTTCCCAAACTACCTGTGGATGTTTCTGTGGATGTTTCTCATGCTTAT	1140
QY	1141	TATGTGGAGTGAATATAGATTCTCACCAATTTAGAAATGGACAGAGATTTGTGTTTT	1200
DB	1141	TATGTGGAGTGAATATAGATTCTCACCAATTTAGAAATGGACAGAGATTTGTGTTTT	1200
QY	1201	ATCTCTGGGTAAATACGTTTTTCTCCAGTTGTATAAGACCTCCCAAGCTATAAAGTC	1260
DB	1201	ATCTCTGGGTAAATACGTTTTTCTCCAGTTGTATAAGACCTCCCAAGCTATAAAGTC	1260
QY	1261	CTATGCAACAAAGAAAATGTCATACATTCCTTAGTCTCATTTATTTTTCATTAGATA	1320
DB	1261	CTATGCAACAAAGAAAATGTCATACATTCCTTAGTCTCATTTATTTTTCATTAGATA	1320
QY	1321	GCCGGTTTTTACTACAACTCAATTAAGATGAACAGAAATGAATGGTGTAGTGACTGTTA	1380
DB	1321	GCCGGTTTTTACTACAACTCAATTAAGATGAACAGAAATGAATGGTGTAGTGACTGTTA	1380
QY	1381	TAAAGAGAGTAATAAGATACATCATTTTGGGCAATTAAGGGAGGAGAGATTTCAG	1440
DB	1381	TAAAGAGAGTAATAAGATACATCATTTTGGGCAATTAAGGGAGGAGAGATTTCAG	1440
QY	1441	CAAAAGTGTCTTCAAGTGGAAAAACAAGTTAAACTTAAAGTGAACCTCTCTTGACAA	1500
DB	1441	CAAAAGTGTCTTCAAGTGGAAAAACAAGTTAAACTTAAAGTGAACCTCTCTTGACAA	1500
QY	1501	GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAAATTTGCTTCTGATA	1560
DB	1501	GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAAATTTGCTTCTGATA	1560
QY	1561	AGCCTGTTTCAATAATCTCTTTTGCAGAGCTCTGCTACTTACCAGAGTCTGCCCTACAGAC	1620
DB	1561	AGCCTGTTTCAATAATCTCTTTTGCAGAGCTCTGCTACTTACCAGAGTCTGCCCTACAGAC	1620
QY	1621	AGAAAGTGCATTTAAACGGTACTTGGAAAAATAGAAAAATGAGAACTATGAAAAATTC	1680
DB	1621	AGAAAGTGCATTTAAACGGTACTTGGAAAAATAGAAAAATGAGAACTATGAAAAATTC	1680
QY	1681	ATGGAAGCAATGGGTAAAGCTTACTTTTGTGAATGCTTCTTAAACAGCAGTACCACTAC	1740
DB	1681	ATGGAAGCAATGGGTAAAGCTTACTTTTGTGAATGCTTCTTAAACAGCAGTACCACTAC	1740
QY	1741	GGCGGAATCAAACTTAAGCTGTTTCACTACCACTCTGGCTTAACCTGCTTCTTGTGT	1800
DB	1741	GGCGGAATCAAACTTAAGCTGTTTCACTACCACTCTGGCTTAACCTGCTTCTTGTGT	1800
QY	1801	CTGCTATTTTGGCTTGCACATTTGCCCTGCACTTTTGAAGAAAGCTCTATAGAGGGA	1860
DB	1801	CTGCTATTTTGGCTTGCACATTTGCCCTGCACTTTTGAAGAAAGCTCTATAGAGGGA	1860
QY	1861	ATCAAGGAAGAAAAACATTTCTGATTTTATTTTGTGCAATAATCTTATGCAATTTAGCT	1920

Db 1861 ATACAAGCAAGAAAAACATTCGATTTATTTGCAATTCGCAATAATCTTATGCAATTAGCT 1920  
Qy 1921 AATTCCAGTAGAGGCATTCAGAGCAAGAAATTTAATAGAAATATATGTAAGGAATATATT 1980  
Db 1921 AATTCCAGTAGAGGCATTCAGAGCAAGAAATTTAATAGAAATATATGTAAGGAATATATT 1980  
Qy 1981 TTGATAAGACTGTTTGAAGAAATACACAAGAGGGAATTCGTCTCCAGTTTTCGAGA 2040  
Db 1981 TTGATAAGACTGTTTGAAGAAATACACAAGAGGGAATTCGTCTCCAGTTTTCGAGA 2040  
Qy 2041 ACACACATGATTGAGTCATTTAAACATGATAGTCTGCTTAAAGCTTTACAAACTGC 2100  
Db 2041 ACACACATGATTGAGTCATTTAAACATGATAGTCTGCTTAAAGCTTTACAAACTGC 2100  
Qy 2101 CTGTATATGGATTAACATACTATCCTAGTTGGATAGTCTGTTGATTAACAGGCTGA 2160  
Db 2101 CTGTATATGGATTAACATACTATCCTAGTTGGATAGTCTGTTGATTAACAGGCTGA 2160  
Qy 2161 ACACAGCTCAGTGAAGAGTGGAGAGTAAGACTCTGAGTCAGAAATTCGGGCTAAGC 2220  
Db 2161 ACACAGCTCAGTGAAGAGTGGAGAGTAAGACTCTGAGTCAGAAATTCGGGCTAAGC 2220  
Qy 2221 TCCTCTCAACTACAGAAAAAGTCACATPAAAAATGCAAAACATGATGTTCTATTGTTTT 2280  
Db 2221 TCCTCTCAACTACAGAAAAAGTCACATPAAAAATGCAAAACATGATGTTCTATTGTTTT 2280  
Qy 2281 CTCTGCTGTGATGTAATGATTAATATTATTATTTTATTTTATTTAGCGGTGAATGATGAAA 2340  
Db 2281 CTCTGCTGTGATGTAATGATTAATATTATTATTTTATTTTATTTAGCGGTGAATGATGAAA 2340  
Qy 2341 GAAAGTTAGGAGCCACGATAATCTGAAGCTCACTATTTCAG 2381  
Db 2341 GAAAGTTAGGAGCCACGATAATCTGAAGCTCACTATTTCAG 2381

RESULT 2  
ID ADL90128 standard; DNA; 336 BP.  
XX  
AC ADL90128;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Chicken intestinal fatty acid binding protein, iFABP, gene, promoter.  
XX  
KW Chicken; ds; intestinal fatty acid binding protein; iFABP;  
KW gut specific promoter; transgenic; promoter.  
XX  
OS Gallus gallus.  
XX  
PN US2003177516-A1.  
XX  
PD 18-SEP-2003.  
XX  
PF 14-MAR-2002; 2002US-00099663.  
XX  
PR 14-MAR-2002; 2002US-00099663.  
XX  
XX (HORS/) HORSEMAN N D.  
PA (PRAT/) PRATT S L.  
XX  
XX Horseman ND, Pratt SL;  
XX  
XX WPI; 2003-898653/82.  
XX  
PT New nucleic acid molecule comprising an isolated avian gut-specific gene  
PT expression control region, useful for regulating heterologous nucleic  
PT acids in transgenic avians, and for generating transgenic birds.  
XX  
PS Claim 1; SEQ ID NO 2; 28pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising an isolated

CC avian gut-specific gene expression control region appearing as  
CC ADL90127(Chicken intestinal fatty acid binding protein, iFABP, gene, 5'  
CC region or ADL90128 (chicken iFABP promoter) or its degenerate variant  
CC Also included are a recombinant DNA molecule comprising an isolate avian  
CC gut-specific gene expression control region operably linked to a nucleic  
CC acid insert encoding a polypeptide, an expression vector that integrates  
CC into a host cell (and comprising the isolated avian gut-specific gene  
CC expression control region), expressing a heterologous polypeptide in a  
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA  
CC molecule, and culturing the transfected cell in a medium suitable for  
CC expression of a heterologous polypeptide under the control of an avian  
CC intestinal fatty acid binding protein (iFABP) or cp35 gene expression  
CC control region encoded by the recombinant DNA molecule), a eukaryotic  
CC cell transformed with the expression vector for its progeny, which  
CC expresses a heterologous polypeptide) and a transgenic avian having a  
CC heterologous polynucleotide sequence comprising the nucleic acid insert.  
CC The nucleic acids are useful for regulating heterologous nucleic acids in  
CC transgenic avians, as probes in nucleic acid hybridisation assays for  
CC detecting the iFABP gene expression control region, and for generating  
CC transgenic birds. The present sequence is the Chicken intestinal fatty  
CC acid binding protein, iFABP, gene, promoter.  
XX  
SQ Sequence 336 BP; 116 A; 62 C; 63 G; 95 T; 0 U; 0 Other;  
  
Query Match 14.1%; Score 336; DB 11; Length 336;  
Best Local Similarity 100.0%; Pred. No. 2.7e-150;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1301 ATTATTATTTCATTAGATAGCCGGTTTTTACTACAACCTCAAATAAGATGAACAGAATG 1360  
Db 1 ATTATTATTTCATTAGATAGCCGGTTTTTACTACAACCTCAAATAAGATGAACAGAATG 60  
  
Qy 1361 AATGGGTTAGTGACTGTTTATAAAGAGAGCTAATAAGATACATCATCTATTGAGGCAA 1420  
Db 61 AATGGGTTAGTGACTGTTTATAAAGAGAGAGTAAAGATACATCATCTATTGAGGCAA 120  
  
Qy 1421 TAAGGGAGGGAGAGATTTCAGCAACACAGTGTCTTACAAGTGGAAAAACAAGTTAACTAAA 1480  
Db 121 TAAGGGAGGGAGAGATTTCAGCAACACAGTGTCTTACAAGTGGAAAAACAAGTTAACTAAA 180  
  
Qy 1481 GTGACCCCTCCTTGACAAGATCAATGCCACAGTTTGAGCTTTAGCCAGGCACATCATCA 1540  
Db 181 GTGACCCCTCCTTGACAAGATCAATGCCACAGTTTGAGCTTTAGCCAGGCACATCATCA 240  
  
Qy 1541 TGTAAATTCGTTTCCTGATAAGCCTGTTCAATAATTTCTTTCARAAGCTCTGCTACTTA 1600  
Db 241 TGTAAATTCGTTTCCTGATAAGCCTGTTCAATAATTTCTTTCARAAGCTCTGCTACTTA 300  
  
Qy 1601 CCAGAAAGTCTGCCCTACAGACAGAAAGATGGCATTTA 1636  
Db 301 CCAGAAAGTCTGCCCTACAGACAGAAAGATGGCATTTA 336  
  
RESULT 3  
AAF22348  
ID AAF22348 standard; cDNA; 1160 BP.  
XX  
AC AAF22348;  
XX  
DT 26-MAR-2001 (first entry)  
XX  
XX Human secreted protein gene 33 SEQ ID NO:43.  
XX  
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmologic; vulnery; gene therapy; neoplasm;  
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system infection; infection;  
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;  
KW skin aging; food additive; preservative; ss.  
XX

OS Homo sapiens.  
 XX WO200061748-A1.  
 FN  
 XX 19-OCT-2000.  
 PD  
 XX  
 XX 06-APR-2000; 2000WO-US008982.  
 PF  
 XX  
 PR 09-APR-1999; 99US-0128696P.  
 PR 14-JAN-2000; 2000US-0176069P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 XX WPI; 2000-638566/61.  
 DR P-PSDB; AAB63081.  
 XX  
 XX New nucleic acid molecules encoding 48 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 XX Claim 1; Page 421; 480pp; English.  
 PS  
 XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049  
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins  
 CC and polypeptides homologous to them. Human secreted proteins have  
 CC activities based on the tissues and cells the genes are expressed in.  
 CC Examples of activities include: immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;  
 CC cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; ophthalmological; and vulnery. The polynucleotides and  
 CC proteins can be used to prevent, treat or ameliorate a medical condition  
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or  
 CC sheep. They are also used in diagnosing a pathological condition or  
 CC susceptibility to a pathological condition. Disorders which are diagnosed  
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
 CC ocular disorders e.g. corneal infection. The polypeptides can also be  
 CC used to aid wound healing and epithelial cell proliferation, to prevent  
 CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities, fat content,  
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other  
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent  
 CC sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;  
 Query Match 1.0%; Score 24; DB 3; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2297 TTGATTATTATTATTTTTTTTTT 2320  
 Db 233 TTGATTATTATTATTTTTTTTTT 256  
 RESULT 4  
 ADC20223  
 ID ADC20223 standard; DNA; 1160 BP.  
 XX  
 AC ADC20223;  
 XX  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX Human secreted protein coding sequence #162.  
 DE  
 XX gene therapy; human; secreted protein; haemopoietic disorder;  
 KW

KW haematological disorder; anaemia; haemophilia; inflammatory disorder;  
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;  
 KW leukaemia; wound healing; epithelial cell proliferation disorder;  
 KW immune disorder; autoimmune disorder; asthmatic disorder;  
 KW cardiovascular disorder; atherosclerosis; myocarditis;  
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;  
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.  
 XX Homo sapiens.  
 OS  
 XX WO200292787-A2.  
 PN  
 XX 21-NOV-2002.  
 PD  
 XX  
 XX 26-MAR-2002; 2002WO-US009257.  
 PF  
 XX  
 PR 27-MAR-2001; 2001US-0278650P.  
 PR 12-SEP-2001; 2001US-00950082.  
 PR 12-SEP-2001; 2001US-00950083.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Rosen CA, Ruben SM;  
 PI  
 XX WPI; 2003-129287/12.  
 DR  
 XX New human secreted proteins and nucleic acid molecules, useful for  
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
 PT preventing or treating hematopoietic or hematologic disorders, e.g.  
 PT anemia or hemophilia.  
 XX  
 PS Claim 1; SEQ ID NO 172; 1512pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequences of human  
 CC secreted proteins. The DNA and protein sequences of the invention are  
 CC useful for detecting, preventing, diagnosing, prognosticating, treating  
 CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia  
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease  
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);  
 CC wound healing and disorders of epithelial cell proliferation; immune  
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);  
 CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);  
 CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);  
 CC and gastrointestinal disorders (e.g. duodenal ulcers and  
 CC gastroenteritis). The present DNA sequence encodes a human secreted  
 CC protein of the invention.  
 XX  
 SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;  
 Query Match 1.0%; Score 24; DB 10; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2297 TTGATTATTATTATTTTTTTTTT 2320  
 Db 233 TTGATTATTATTATTTTTTTTTT 256  
 RESULT 5  
 ABT16829  
 ID ABT16829 standard; DNA; 1160 BP.  
 XX  
 AC ABT16829;  
 XX  
 XX 03-APR-2003 (first entry)  
 DT  
 XX Human secreted protein gene sequence - SEQ ID No 78.  
 DE  
 XX Human; gene; ds; protein therapy; immediate hypersensitivity disease;  
 KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;  
 KW hay fever; allergic conjunctivitis; allergic rhinitis;  
 KW binding partner identification; chromosome identification;  
 KW radiation hybrid mapping; long-range restriction mapping.



```

XX OS Homo sapiens.
XX XN WO200277188-A2.
XX PD 03-OCT-2002.
XX XX
XX PF 26-MAR-2002; 2002WO-US009239.
XX XX
XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2003-175010/17.
XX XX
XX PT Use of human secreted proteins and nucleic acids for preparing a
XX PT diagnostic or pharmaceutical composition for diagnosing or treating
XX PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
XX PT conjunctivitis or rhinitis.
XX PS Claim 7; Page 601-602; 823pp; English.
XX XX
XX CC The invention comprises the amino acid and coding sequences of human
XX CC secreted proteins. The DNA and protein sequences of the invention are
XX CC useful for the diagnosis and treatment of allergic disorders, asthmatic
XX CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
XX CC allergic conjunctivitis and allergic rhinitis). The proteins of the
XX CC invention are also useful for identifying a binding partner. The nucleic
XX CC acids of the invention are also useful for chromosome identification.
XX CC radiation hybrid mapping or long-range restriction mapping. The present
XX CC DNA sequence encodes a human secreted protein of the invention
XX SQ
XX SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
Query Match 1.0%; Score 24; DB 10; Length 1160;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2297 TTGATTATTATTATTTTTTTTTT 2320
Db 233 TTGATTATTATTATTTTTTTTTT 256
RESULT 6
ABZ67143
ID ABZ67143 standard; cDNA; 1160 BP.
XX AC ABZ67143;
XX XX
XX DT 26-MAR-2003 (first entry)
XX DE Human secreted protein encoding cDNA SEQ ID NO 263.
XX XX
XX KW Human; secreted protein; neurotropic; neuroprotective; cytostatic;
XX KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
XX KW vulnary; antibacterial; antiparkinsonian; antiskilling; antianaemic;
XX KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
XX KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
XX KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
XX KW cardiovascular disorder; neurological disease; nephrotropic;
XX KW gene therapy; gene; ds.
XX OS Homo sapiens.
XX XX
XX PN WO200277186-A2.
XX XX
XX PD 03-OCT-2002.
XX XX
XX PF 26-MAR-2002; 2002WO-US009188.

```

```

XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2003-040583/03.
XX DR P-PSDB; ABP99722.
XX XX
XX PT New human secreted proteins encoded by genes contained in cDNA clones
XX PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
XX PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
XX PT West Nile fever.
XX PS Claim 7; Page 1308; 2423pp; English.
XX XX
XX CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
XX CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
XX CC treating or ameliorating medical conditions e.g. by protein or gene
XX CC therapy. The genes are isolated from a range of human tissues disclosed
XX CC in the specification. The nucleic acids, proteins, antibodies and
XX CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
XX CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
XX CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
XX CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
XX CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
XX CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
XX CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
XX CC bacterial, fungal and parasitic infections
XX SQ
XX SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
Query Match 1.0%; Score 24; DB 10; Length 1160;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2297 TTGATTATTATTATTTTTTTTTT 2320
Db 233 TTGATTATTATTATTTTTTTTTT 256
RESULT 7
AAS46723
ID AAS46723 standard; DNA; 6130 BP.
XX AC AAS46723;
XX XX
XX DT 18-DEC-2001 (first entry)
XX DE Tumour suppressor gene derived chemically modified sequence #446.
XX XX
XX KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX KW cytosine methylation; ds.
XX OS Homo sapiens.
XX PN WO200168912-A2.
XX XX
XX PD 20-SEP-2001.
XX XX
XX PF 15-MAR-2001; 2001WO-EP002955.
XX XX
XX PR 15-MAR-2000; 2000DE-01013847.
XX PR 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.

```

```

XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2001-602752/68.
XX PT Fragments of chemically modified genes associated with tumor suppressor
XX PT genes and oncogenes, useful in designing primers and probes for analyzing
XX PT diseases associated with cytosine methylation state e.g. cancer.
XX PS Claim 1; SEQ ID NO 446; 27bp; English.
XX CC The invention relates to a nucleic acid comprising a sequence of 18
XX CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX CC bisulphite, of genes associated with tumour suppression and oncogenes
XX CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX CC 500 are missing from the sequence listing) sequences (Ss) and sequences
XX CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
XX CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX CC probes for detecting the cytosine methylation state and/or single
XX CC nucleotide polymorphisms and also to be used in an array for analysing
XX CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX CC probes can also be used in a method for ascertaining genetic and/or
XX CC epigenetic parameters for the diagnosis and/or therapy of existing
XX CC diseases or the predisposition to specific diseases, by analysing
XX CC cytosine methylations. The parameters may be compared to another set of
XX CC genetic and/or epigenetic parameters, the differences serving as basis
XX CC for diagnosis and/or prognosis events which are disadvantageous to
XX CC patients. The present sequence is one of the 533 genomic sequences
XX CC derived from tumour suppressor genes and oncogenes. Sequences with even
XX CC numbered Seq ID numbers are the complementary sequence of the
XX CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
XX CC 535, except for those whose partner sequence is missing). Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 6130 BP; 1568 A; 208 C; 1518 G; 2835 T; 0 U; 1 Other;

Query Match 1.0%; Score 24; DB 4; Length 6130;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2298 TGATTATTATTATTTTATTTTATTA 2321
Db 5133 TGATTATTATTATTTTATTTTATTA 5156

RESULT 8
AAS61411
ID AAS61411 standard; DNA; 6130 BP.
XX AC AAS61411;
XX DT 29-JAN-2002 (first entry)
XX DE Human gene regulation-associated gene oligonucleotide #366.
XX KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
XX KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
XX KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
XX KW renal disease; Preclampsia; cardiac allograft vascular disease;
XX KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
XX KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
XX KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX OS Homo sapiens.
XX PN WO200177375-A2.
XX PD 18-OCT-2001.
XX PF
XX PP
XX PR

```

```

PF 06-APR-2001; 2001WO-EP003968.
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017470/02.
XX
XX New nucleic acid sequences from chemically modified genes associated with
XX PT gene regulation, useful for analyzing cytosine methylations for diagnosis
XX PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX PS Disclosure; SEQ ID NO 374; 26pp; English.
XX CC The invention relates to 224 nucleic acid sequences comprising at least
XX CC 18 bases of a chemically pretreated gene associated with gene regulation
XX CC selected from 43 known genes (or complementary sequences). The chemical
XX CC pretreatment converts cytosine bases unmethylated at the 5-position to
XX CC uracil or another base with hybridisation behaviour dissimilar to
XX CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
XX CC oligomers (or sets/arrays) and method are useful in the diagnosis of
XX CC diseases (or predisposition to diseases) associated with gene regulation
XX CC and in therapy of such diseases, by enabling analysis of the cytosine
XX CC methylation patterns of such genes, kits are provided. They are
XX CC especially useful in diagnosis and therapy of e.g. severe combined
XX CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
XX CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
XX CC syndrome, renal disease, preclampsia, graft versus-host disease. The
XX CC present sequence is a sequence included in the sequence data for this
XX CC specification and is associated with the human gene regulation-associated
XX CC genes. Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 6130 BP; 1568 A; 208 C; 1518 G; 2835 T; 0 U; 1 Other;

Query Match 1.0%; Score 24; DB 6; Length 6130;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2298 TGATTATTATTATTTTATTTTATTA 2321
Db 5133 TGATTATTATTATTTTATTTTATTA 5156

RESULT 9
AAS46388
ID AAS46388 standard; DNA; 7133 BP.
XX AC AAS46388;
XX DT 18-DEC-2001 (first entry)
XX DE Tumour suppressor gene derived chemically modified sequence #110.
XX KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX KW cytosine methylation; ds.
XX OS Homo sapiens.
XX PN WO200168912-A2.
XX PD 20-SEP-2001.
XX PF 15-MAR-2001; 2001WO-EP002955.
XX PP
XX PR 15-MAR-2000; 2000DE-01013847.

```

PR 06-APR-2000; 2000DE-01019058.  
 PR 07-APR-2000; 2000DE-01019173.  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 XX (EPiG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX  
 XX WPI; 2001-602752/68.  
 XX  
 XX Fragments of chemically modified genes associated with tumor suppressor  
 PT genes and oncogenes, useful in designing primers and probes for analyzing  
 PT diseases associated with cytosine methylation state e.g. cancer.  
 XX  
 XX Claim 1; SEQ ID NO 110; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and oncogenes  
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and  
 CC 500 are missing from the sequence listing) sequences (Ss) and sequences  
 CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-  
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of  
 CC probes for detecting the cytosine methylation state and/or single  
 CC nucleotide polymorphisms and also to be used in an array for analysing  
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
 CC probes can also be used in a method for ascertaining genetic and/or  
 CC epigenetic parameters for the diagnosis and/or therapy of existing  
 CC diseases or the predisposition to specific diseases, by analysing  
 CC cytosine methylations. The parameters may be compared to another set of  
 CC genetic and/or epigenetic parameters, the differences serving as basis  
 CC for diagnosis and/or prognosis events which are disadvantageous to  
 CC patients. The present sequence is one of the 533 genomic sequences  
 CC derived from tumour suppressor genes and oncogenes. Sequences with even  
 CC numbered seq ID numbers are the complementary sequence of the  
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
 CC 535, except for those whose partner sequence is missing). Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 7133 BP; 2192 A; 61 C; 1242 G; 3638 T; 0 U; 0 Other;  
 Query Match 1.0%; Score 24; DB 4; Length 7133;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2300 ATTATTATTATTTTATTTTATTTAGG 2323  
 |||||  
 Db 4360 ATTATTATTATTTTATTTTATTTAGG 4383  
 RESULT 10  
 ABL19488/c  
 ID ABL19488 standard; DNA; 22901 BP.  
 XX  
 AC ABL19488;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9937.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI; 2001-656860/75.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 XX Claim 1; SEQ ID NO 9937; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 22901 BP; 7612 A; 4059 C; 3904 G; 7326 T; 0 U; 0 Other;  
 Query Match 1.0%; Score 24; DB 4; Length 22901;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2297 TTGATTATTATTATTTTATTTT 2320  
 |||||  
 Db 4254 TTGATTATTATTATTTTATTTT 4231  
 RESULT 11  
 AAD58279/c  
 ID AAD58279 standard; DNA; 226475 BP.  
 XX  
 AC AAD58279;  
 XX  
 XX 20-NOV-2003 (first entry)  
 DT  
 DE Human tumour suppressor gene, Lmt reverse complement DNA.  
 XX  
 KW Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003066869-A1.  
 PN  
 XX 14-AUG-2003.  
 PD  
 XX 07-FEB-2003; 2003WO-AU000126.  
 PF  
 XX 07-FEB-2002; 2002AU-00000371.  
 PR  
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.  
 XX  
 XX Cook WD, Mccaw BJ;  
 PI  
 XX WPI; 2003-646311/61.  
 XX  
 XX New nucleic acid molecule, useful for screening a subject for the  
 PT presence of an aberration in a gene encoding an LMT.  
 PT  
 XX Claim 10; Page 233-299; 373pp; English.  
 PS  
 CC The invention relates to novel tumour suppressor gene, referred to as  
 CC Lmt. The invention also relates to the field of cancer therapy and cancer

CC diagnostics. The nucleic acid molecule is useful for screening a subject  
CC for the presence of an aberration in a gene encoding an LMT. The present  
CC sequence is human Lmt reverse complement DNA

XX SQ Sequence 226475 BP; 61024 A; 41761 C; 40916 G; 57494 T; 0 U; 25280 Other;  
Query Match 1.0%; Score 24; DB 9; Length 226475;

Best Local Similarity 100.0%; Pred. No. 0.98;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2297 TTGATTATTATTATTATTTT 2320

Db 36606 TTGATTATTATTATTATTTT 36583

RESULT 12

ADL90141/C

ID ADL90141 standard; DNA; 23 BP.

XX AC ADL90141;

DT 20-MAY-2004 (first entry)

DE Chicken iFABP, cis promoter element primer #1.

XX Chicken; ss; intestinal fatty acid binding protein; iFABP;  
KW gut specific promoter; transgenic; PCR; primer.

XX Gallus gallus.

OS US2003177516-A1.

PN 18-SEP-2003.

XX 14-MAR-2002; 2002US-00099663.

PF 14-MAR-2002; 2002US-00099663.

PR (HORS/) HORSEMAN N D.  
PA (PRATT/) PRATT S L.

XX Horseman ND, Pratt SL;

PI WPI; 2003-898653/82.

DR New nucleic acid molecule comprising an isolated avian gut-specific gene  
XX expression control region, useful for regulating heterologous nucleic  
PT acids in transgenic avians, and for generating transgenic birds.

PS Example 5; SEQ ID NO 15; 28pp; English.

XX The invention relates to an isolated nucleic acid comprising an isolated  
CC avian gut-specific gene expression control region appearing as  
CC ADL90127(Chicken intestinal fatty acid binding protein, iFABP, gene, 5'  
CC region or ADL90128 (chicken iFABP promoter) or its degenerate variant.  
CC Also included are a recombinant DNA molecule comprising an isolate avian  
CC gut-specific gene expression control region operably linked to a nucleic  
CC acid insert encoding a polypeptide, an expression vector that integrates  
CC into a host cell (and comprising the isolated avian gut-specific gene  
CC expression control region), expressing a heterologous polypeptide in a  
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA  
CC molecule, and culturing the transfected cell in a medium suitable for  
CC expression of a heterologous polypeptide under the control of an avian  
CC intestinal fatty acid binding protein (iFABP) or cp35 gene expression  
CC control region encoded by the recombinant DNA molecule), a eukaryotic  
CC cell transformed with the expression vector (or its progeny, which  
CC expresses a heterologous polypeptide) and a transgenic avian having a  
CC heterologous polynucleotide sequence comprising the nucleic acid insert.  
CC The nucleic acids are useful for regulating heterologous nucleic acids in  
CC transgenic avians, as probes for nucleic acid hybridisation assays for  
CC detecting the iFABP gene expression control region, and for generating  
CC transgenic birds. The present sequence is a PCR primer used to isolate  
CC the Chicken iFABP gene cis promoter element.

XX

SQ Sequence 23 BP; 3 A; 6 C; 5 G; 9 T; 0 U; 0 Other;

Query Match

Best Local Similarity 1.0%; Score 23; DB 11; Length 23;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1604 GAAGTCTGCTACACAGAGAAG 1626

Db 23 GAAGTCTGCTACACAGAGAAG 1

RESULT 13

ABX39142

ID ABX39142 standard; cDNA; 472 BP.

XX AC ABX39142;

DT 20-FEB-2003 (first entry)

DE Bovine EST associated with lactation/muscle/fat deposition #4307.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;

KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW Gene analysis; cattle breeding.

XX Bos Taurus.

OS US2002137139-A1.

PN 26-SEP-2002.

XX 24-SEP-2001; 2001US-00960352.

PF 12-JAN-1999; 99US-0115707P.

PR 11-JAN-2000; 2000US-00480902.

XX (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

PI WPI; 2003-110599/10.

DR New nucleic acid associated with lactation, and muscle and fat  
XX deposition, useful for genome mapping, gene identification and analysis,  
PT cattle breeding, or for genetically improving cattle.

PS Claim 2; SEQ ID NO 4307; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived from  
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second  
CC nucleic acid molecule comprising any of 1512 nucleotide sequences,  
CC appearing as ABX3486-ABX4947, or complements of them. Also included are  
CC (1) a transformed cell having a nucleic acid comprising an LMFD nucleic  
CC acid linked to a promoter and a 3' non-translated sequence that  
CC functions in the cell to cause termination of transcription and addition  
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
CC (2) determining a level or pattern of a molecule in a bovine cell or  
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
CC of the 1512 nucleic acid sequences or its complement or fragment) with a  
CC complementary nucleic acid molecule obtained from the bovine cell or  
CC tissue, where hybridisation between the marker nucleic acid and the  
CC complementary nucleic acid permits the detection of the molecule; and (b)  
CC detecting the level or pattern of the complementary nucleic acid, where  
CC the detection of the complementary nucleic acid is predictive of the  
CC level or pattern of the molecule. The LMFD nucleic acid is used for  
CC determining a level or pattern of a molecule in a bovine cell or tissue.  
CC It is useful for genome mapping, gene identification and analysis, cattle  
CC breeding, preparation of constructs for use in cattle gene expression, or

CC for genetically improving cattle. The present sequence is one of the  
CC 15112 bovine IMFED EST (expressed sequence tag) nucleic acids. Note: The  
CC present sequence was not shown in the specification but was obtained in  
CC electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137139

XX  
SQ Sequence 472 BP; 109 A; 91 C; 109 G; 163 T; 0 U; 0 Other;

Query Match 1.0%; Score 23; DB 8; Length 472;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTTTAG 2322  
|||||  
Db 441 ATTATTATTATTTTTTTTTTTAG 463

## RESULT 14

ABV96362  
ID ABV96362 standard; cDNA; 503 BP.

XX  
AC ABV96362;

XX  
DT 14-JAN-2003 (first entry)

XX  
DE Human pancreatic cancer expressed cDNA SEQ ID NO 1770.

XX  
KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;  
KW cytostatic; tumour; gene; ss.

XX  
OS Homo sapiens.

XX  
FN WC200260317-A2.

XX  
PD 08-AUG-2002.

XX  
PF 30-JAN-2002; 2002WO-US002781.

XX  
PR 30-JAN-2001; 2001US-0265305P.

XX  
PR 31-JAN-2001; 2001US-0265682P.

XX  
PR 09-FEB-2001; 2001US-0267588P.

XX  
PR 21-MAR-2001; 2001US-0278651P.

XX  
PR 28-APR-2001; 2001US-0287112P.

XX  
PR 16-MAY-2001; 2001US-0291631P.

XX  
PR 12-JUL-2001; 2001US-0305484P.

XX  
PR 20-AUG-2001; 2001US-0313999P.

XX  
PR 27-NOV-2001; 2001US-0333626P.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;

XX  
PI WPI; 2002-627435/67.

XX  
PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for  
PT diagnosing, preventing and/or treating cancer, particularly pancreatic  
PT cancer.

XX  
PS Claim 1; SEQ ID NO 1770; 300pp + Sequence Listing; English.

XX  
CC The invention relates to an isolated polynucleotide (I) comprising: (a)  
CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)  
CC complements of (a); (c) sequences consisting of at least 20 contiguous  
CC residues of (a); (d) sequences that hybridize to (a) under moderately  
CC stringent conditions; (e) sequences having at least 75% or 90% identity  
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-  
CC ABP68637) encoded by (i) and oligonucleotide can be used to detect cancer  
CC in a patient and compositions comprising polypeptides, polynucleotides,  
CC antibodies, fusion proteins, T cell populations and antigen presenting  
CC cells expressing the polypeptide are useful in treating pancreatic cancer  
CC and stimulating an immune response. The polynucleotides can be used as  
CC probes or primers for nucleic acid hybridisation, in the design and  
CC preparation of ribozyme molecules for inhibiting expression of the tumour

CC polypeptides and proteins in the tumour cells, in vaccines and for gene  
CC therapy. Note: The sequence data for this patent did not form part of the  
CC printed specification but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 503 BP; 152 A; 91 C; 101 G; 155 T; 0 U; 4 Other;

Query Match 1.0%; Score 23; DB 6; Length 503;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTTTAG 2322  
|||||  
Db 63 ATTATTATTATTTTTTTTTTTAG 85

## RESULT 15

ADC20846  
ID ADC20846 standard; DNA; 1112 BP.

XX  
AC ADC20846;

XX  
DT 18-DEC-2003 (first entry)

XX  
DE Human secreted protein-related DNA sequence #264.

XX  
KW gene therapy; human; secreted protein; haemopoietic disorder;  
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;  
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;  
KW leukaemia; wound healing; epithelial cell proliferation disorder;  
KW immune disorder; autoimmune disorder; asthmatic disorder;  
KW cardiovascular disorder; atherosclerosis; myocarditis;  
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;  
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.

XX  
OS Homo sapiens.

XX  
FN WC200292787-A2.

XX  
PD 21-NOV-2002.

XX  
PF 26-MAR-2002; 2002WO-US0009257.

XX  
PR 27-MAR-2001; 2001US-0278650P.

XX  
PR 12-SEP-2001; 2001US-00950082.

XX  
PR 12-SEP-2001; 2001US-00950083.

XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI Rosen CA, Ruben SM;

XX  
PI WPI; 2003-129287/12.

XX  
PT New human secreted proteins and nucleic acid molecules, useful for  
PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
PT preventing or treating hematopoietic or hematologic disorders, e.g.  
PT anemia or hemophilia.

XX  
PS Disclosure; SEQ ID NO 800; 1512pp; English.

XX  
CC The invention comprises the amino acid and coding sequences of human  
CC secreted proteins. The DNA and protein sequences of the invention are  
CC useful for detecting, preventing, diagnosing, prognosticating, treating  
CC or ameliorating; hematopoietic or haematological disorders (e.g. anaemia  
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease  
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);  
CC wound healing and disorders of epithelial cell proliferation; immune  
CC disorders (e.g. autoimmune disorders and asthmatic disorders);  
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);  
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);  
CC and gastrointestinal disorders (e.g. duodenal ulcers and  
CC gastroenteritis). The present DNA sequence was used in the  
CC exemplification of the invention.

```

XX
SQ  Sequence 1112 BP; 267 A; 221 C; 178 G; 446 T; 0 U; 0 Other;
    Query Match      1.0%; Score 23; DB 10; Length 1112;
    Best Local Similarity 100.0%; Pred. No. 3.3;
    Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy  2297 TTGATTATTATTATTATTATT 2319
Db  233 TTGATTATTATTATTATTATT 255

```

Search completed: January 14, 2005, 19:52:32  
Job time : 1138 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 19:07:22 ; Search time 203 Seconds  
(without alignments)  
8336.888 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agcttctgcgcagaaagg.....atctgaagctcactattcag 2381

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 35539441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	0.9	1299	3	US-08-793-035-3
C 2	22	0.9	6124	3	US-08-213-419B-3
C 3	22	0.9	10619	4	US-10-204-708-4
C 4	21	0.9	398	4	US-09-513-999C-8929
C 5	21	0.9	810	4	US-09-270-767-26506
C 6	21	0.9	1072	4	US-09-270-767-11001
C 7	21	0.9	1701	3	US-09-078-294-9
C 8	21	0.9	1751	4	US-09-270-767-6519
C 9	21	0.9	1751	4	US-09-270-767-21801
C 10	21	0.9	3001	4	US-09-539-333D-211
C 11	21	0.9	3286	3	US-09-211-417-2
C 12	21	0.9	12847	1	US-08-550-715-1
C 13	21	0.9	43950	3	US-09-735-934A-3
C 14	21	0.9	43950	4	US-10-060-332-3
C 15	21	0.9	43950	4	US-10-339-657-3
C 16	21	0.9	72604	3	US-09-268-992-7
C 17	21	0.9	72604	3	US-09-657-474-7
C 18	21	0.9	80246	3	US-09-078-294-4
C 19	21	0.9	80595	3	US-09-078-294-3
C 20	21	0.9	81001	4	US-09-750-580-1
C 21	21	0.9	162450	3	US-09-345-882-1
C 22	20	0.8	245	4	US-09-513-999C-35624
C 23	20	0.8	376	4	US-09-270-767-27581
C 24	20	0.8	503	4	US-09-270-767-7474
C 25	20	0.8	503	4	US-09-270-767-22756
C 26	20	0.8	764	4	US-09-270-767-7079
C 27	20	0.8	764	4	US-09-270-767-22361

28	20	0.8	1597	4	US-09-270-767-11916	Sequence 11916, A
29	20	0.8	1834	1	US-08-297-633A-1	Sequence 1, Appli
30	20	0.8	1834	1	US-08-485-721-8	Sequence 8, Appli
31	20	0.8	1834	2	US-08-392-935-8	Sequence 8, Appli
32	20	0.8	1834	5	PCT-US93-08325-1	Sequence 1, Appli
33	20	0.8	1834	5	PCT-US93-08326-8	Sequence 8, Appli
34	20	0.8	2417	1	US-08-011-398B-1	Sequence 1, Appli
35	20	0.8	2417	1	US-08-464-051-1	Sequence 1, Appli
36	20	0.8	2417	2	US-08-462-498-1	Sequence 1, Appli
37	20	0.8	2417	3	US-08-554-385-2	Sequence 2, Appli
C 38	20	0.8	2592	4	US-09-620-312B-301	Sequence 301, App
39	20	0.8	3254	3	US-08-965-903B-1	Sequence 1, Appli
C 40	20	0.8	3973	2	US-08-602-093-6	Sequence 1, Appli
41	20	0.8	5219	4	US-10-204-708-52	Sequence 52, Appl
42	20	0.8	5562	4	US-10-204-708-63	Sequence 63, Appl
43	20	0.8	6801	4	US-10-204-708-62	Sequence 62, Appl
44	20	0.8	19513	4	US-10-204-708-40	Sequence 40, Appl
C 45	20	0.8	41684	4	US-09-536-059-1	Sequence 1, Appli

RESULT 1  
US-08-793-035-3/c  
; Sequence 3, Application US/08793035  
; Patent No. 6011201  
; GENERAL INFORMATION:  
; APPLICANT: Slabas, Antoni R.  
; APPLICANT: White, Andrew  
; APPLICANT: Chase, Dianne  
; APPLICANT: Elborough, Keiran  
; APPLICANT: Featem, Phillip A.  
; TITLE OF INVENTION: B-Ketoacyl ACP Reductase Genes From  
; TITLE OF INVENTION: Brassica Napus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: US  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,035  
; FILING DATE: 28-JUL-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9414622.2  
; FILING DATE: 20-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB95/01678  
; FILING DATE: 17-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kammerer, Patricia A.  
; REGISTRATION NUMBER: 29,775  
; REFERENCE/DOCKET NUMBER: MOBT:132  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713.787.1400  
; TELEFAX: 713.787.1440  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-793-035-3

ALIGNMENTS

```
Query Match          0.9%; Score 22; DB 3; Length 1299;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2299 GATTATTATTATTATTTT 2320
Db      687 GATTATTATTATTATTTT 666

RESULT 2
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE OF INVENTION: AND USE THEREFOR
; FILE REFERENCE: J11-002CNCPC
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match          0.9%; Score 22; DB 3; Length 6124;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTT 2321
Db      5988 ATTATTATTATTTT 6009

RESULT 3
US-10-204-708-4
; Sequence 4, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 4

Query Match          0.9%; Score 22; DB 3; Length 1299;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; LENGTH: 10619
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-4

Query Match          0.9%; Score 22; DB 4; Length 10619;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2301 TTATTATTATTTT 2322
Db      2915 TTATTATTATTTT 2936

RESULT 4
US-09-513-999C-8929
; Sequence 8929, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 8929
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 61
; OTHER INFORMATION: h=a or c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 90
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 267
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 292
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 317
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 355
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 396
; OTHER INFORMATION: s=g or c
; OTHER INFORMATION: 8929
US-09-513-999C-8929

Query Match          0.9%; Score 21; DB 4; Length 398;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTT 2320
Db      2915 TTATTATTATTTT 2936
```



```
Db      238 ATTATTATTATTATTTTTTTTTT 258

RESULT 5
US-09-270-767-26506/c
; Sequence 26506, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26506
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-26506

Query Match          0.9%; Score 21; DB 4; Length 810;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2300 ATTATTATTATTTTTTTTTT 2320
      |||
Db      595 ATTATTATTATTTTTTTTTT 575

RESULT 6
US-09-270-767-11001/c
; Sequence 11001, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11001
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11001

Query Match          0.9%; Score 21; DB 4; Length 1072;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2300 ATTATTATTATTTTTTTTTT 2320
      |||
Db      595 ATTATTATTATTTTTTTTTT 575

RESULT 7
US-09-270-294-9
; Sequence 9, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1701

; TYPE: DNA
; ORGANISM: BAC-F2 contig 5
US-09-078-294-9

Query Match          0.9%; Score 21; DB 3; Length 1701;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2300 ATTATTATTATTTTTTTTTT 2320
      |||
Db      1097 ATTATTATTATTTTTTTTTT 1117

RESULT 8
US-09-270-767-6519/c
; Sequence 6519, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6519
; LENGTH: 1751
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6519

Query Match          0.9%; Score 21; DB 4; Length 1751;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2300 ATTATTATTATTTTTTTTTT 2320
      |||
Db      816 ATTATTATTATTTTTTTTTT 796

RESULT 9
US-09-270-767-21801/c
; Sequence 21801, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21801
; LENGTH: 1751
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-21801

Query Match          0.9%; Score 21; DB 4; Length 1751;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2300 ATTATTATTATTTTTTTTTT 2320
      |||
Db      816 ATTATTATTATTTTTTTTTT 796

RESULT 10
US-09-539-333D-211/c
; Sequence 211, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
```

APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueret, Lydie  
APPLICANT: Bhain, Bernard  
APPLICANT: Esioux, Laurent  
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS  
FILE REFERENCE: GENSET.047AUS  
CURRENT APPLICATION NUMBER: US 09/539,333D  
CURRENT FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/126,903  
PRIOR FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: US 60/131,971  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/132,065  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/143,928  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: US 60/145,915  
PRIOR FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: US 60/146,453  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/146,452  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/162,288  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: US 09/416,384  
PRIOR FILING DATE: 1999-10-12  
NUMBER OF SEQ ID NOS: 231  
SOFTWARE: Patent.pm  
SEQ ID NO 211  
LENGTH: 3001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 1501  
OTHER INFORMATION: 99-26781-25 : polymorphic base G or T  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1482..1500  
OTHER INFORMATION: 99-26781-25.mis1  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1502..1521  
OTHER INFORMATION: 99-26781-25.mis2, complement  
FEATURE:  
NAME/KEY: primer bind  
LOCATION: 1477..1497  
OTHER INFORMATION: upstream amplification primer  
FEATURE:  
NAME/KEY: primer bind  
LOCATION: 1905..1925  
OTHER INFORMATION: downstream amplification primer, complement  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1489..1513  
OTHER INFORMATION: 99-26781-25 probe  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 21,274..275  
OTHER INFORMATION: n=a, g, c or t  
US-09-539-333D-211

Query Match 0.9%; Score 21; DB 4; Length 3001;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2300 ATTATTATTATTTTTTTT 2320  
|||||  
Db 448 ATTATTATTATTTTTTTT 428  
  
RESULT 11

US-09-211-417-2  
Sequence 2, Application US/09211417A  
Patent No. 6177254  
GENERAL INFORMATION:  
APPLICANT: Rattner, Jerome B  
APPLICANT: Whitehead, Clark M  
TITLE OF INVENTION: NUCLEOLUS AUTOANTIGENIC MARKER FOR SYSTEMIC LUPUS  
TITLE OF INVENTION: ERHYEMATOSUS  
TITLE OF INVENTION: GenBank  
TITLE OF INVENTION: Accession Number U86751  
FILE REFERENCE: UCCI  
CURRENT APPLICATION NUMBER: US/09/211,417A  
CURRENT FILING DATE: 1998-12-15  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 3286  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Nucleic Acid Sequence of ASE-1  
Patent No. 6177254  
US-09-211-417-2  
  
Query Match 0.9%; Score 21; DB 3; Length 3286;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2300 ATTATTATTATTTTTTTT 2320  
|||||  
Db 2406 ATTATTATTATTTTTTTT 2426  
  
RESULT 12  
US-08-550-715-1  
Sequence 1, Application US/08550715  
Patent No. 5750345  
GENERAL INFORMATION:  
APPLICANT: Bowie, Lemuel J.  
TITLE OF INVENTION: Human  $\alpha$ -Thalassemia Mutations as a Predictor of  
TITLE OF INVENTION: Blood-Related Disorders  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/550,715  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28493/32834  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12847 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA

```
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(6703..6797, 6915..7119, 7262..7387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(10514..10608, 10736..10930, 11080..11205)
US-08-550-715-1

Query Match          0.9%; Score 21; DB 1; Length 12847;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTT 2320
|||||
Db 5010 ATTATTATTATTTTTTTTTT 5030

RESULT 13
US-09-735-934A-3/c
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match          0.9%; Score 21; DB 3; Length 43950;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTT 2320
|||||
Db 29144 ATTATTATTATTTTTTTTTT 29124

RESULT 14
US-10-060-332-3/c
; Sequence 3, Application US/10060332
; Patent No. 6528294
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851DIV
; CURRENT APPLICATION NUMBER: US/10/060,332
; CURRENT FILING DATE: 2002-02-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-060-332-3

Query Match          0.9%; Score 21; DB 4; Length 43950;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTT 2320
|||||
Db 29144 ATTATTATTATTTTTTTTTT 29124

RESULT 15
US-10-339-657-3/c
; Sequence 3, Application US/10339657
; Patent No. 6780626
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851DIV-2
; CURRENT APPLICATION NUMBER: US/10/339,657
; CURRENT FILING DATE: 2003-01-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-339-657-3

Query Match          0.9%; Score 21; DB 4; Length 43950;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTT 2320
|||||
Db 29144 ATTATTATTATTTTTTTTTT 29124

Search completed: January 15, 2005, 00:47:08
Job time : 207 secs
```

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 22:40:37 ; Search time 1275 Seconds  
(without alignments)  
10730.165 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agcttcctacacac

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4300275 seqs, 2872944193 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8600550

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubnpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubnpna/FACT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubnpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubnpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubnpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubnpna/FACTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubnpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubnpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubnpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubnpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubnpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubnpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubnpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubnpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubnpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubnpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubnpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubnpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubnpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubnpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubnpna/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID		Description
	Score	Match	Length	DB		
1	2381	100.0	2381	15	US-10-099-663-1	Sequence 1, Appli
2	336	14.1	336	15	US-10-099-663-2	Sequence 2, Appli
3	24	1.0	6130	16	US-10-221-613-374	Sequence 374, App
4	24	1.0	6130	16	US-10-221-714A-448	Sequence 448, App
5	24	1.0	7133	16	US-10-221-714A-110	Sequence 110, App
6	24	1.0	3673778	15	US-10-312-841-1	Sequence 1, Appli
7	23	1.0	23	23	US-10-099-663-15	Sequence 15, Appli
8	23	1.0	472	9	US-09-960-352-4307	Sequence 4307, Ap
9	23	1.0	503	14	US-10-060-036-1770	Sequence 1770, App
10	23	1.0	6621	15	US-10-311-455-885	Sequence 885, App
11	23	1.0	10139	9	US-09-764-846-335	Sequence 335, App
12	23	1.0	10139	14	US-10-091-483-335	Sequence 335, App

C	13	23	1.0	107543	17	US-10-322-281-706	Sequence 706, Appl
	14	23	1.0	169659	17	US-10-322-696-707	Sequence 70, Appl
C	15	23	1.0	175737	18	US-10-723-860-1753	Sequence 1753, Ap
C	16	22	0.9	22	15	US-10-099-663-11	Sequence 11, Appl
C	17	22	0.9	22	15	US-10-099-663-13	Sequence 13, Appl
C	18	22	0.9	22	15	US-10-099-663-17	Sequence 17, Appl
	19	22	0.9	201	17	US-10-741-601-24770	Sequence 24770, A
	20	22	0.9	636	13	US-10-027-632-284290	Sequence 284290, S
	21	22	0.9	636	15	US-10-027-632-284290	Sequence 284290, S
	22	22	0.9	830	13	US-10-027-632-31605	Sequence 31605, A
	23	22	0.9	830	15	US-10-027-632-31605	Sequence 31605, A
C	24	22	0.9	830	17	US-10-767-795-2033	Sequence 2033, Ap
	25	22	0.9	1159	16	US-10-424-599-48390	Sequence 48390, A
	26	22	0.9	2000	9	US-09-938-842A-4413	Sequence 4413, Ap
	27	22	0.9	2000	11	US-09-938-842A-4413	Sequence 4413, Ap
	28	22	0.9	4274	15	US-10-311-455-2039	Sequence 2039, Ap
	29	22	0.9	5388	16	US-10-321-613-199	Sequence 199, App
	30	22	0.9	6076	16	US-10-221-714A-385	Sequence 385, App
C	31	22	0.9	6800	15	US-10-311-455-468	Sequence 468, App
	32	22	0.9	10619	14	US-10-239-676-2	Sequence 2, Appli
	33	22	0.9	10619	15	US-10-204-708-4	Sequence 4, Appli
	34	22	0.9	10619	15	US-10-311-455-44	Sequence 44, Appl
	35	22	0.9	10619	15	US-10-240-453-2	Sequence 2, Appli
	36	22	0.9	10619	16	US-10-240-589C-2	Sequence 2, Appli
C	37	22	0.9	15224	16	US-10-721-613-223	Sequence 223, App
	38	22	0.9	193691	18	US-10-719-993-6768	Sequence 6768, Ap
C	39	22	0.9	321491	13	US-10-087-192-532	Sequence 532, App
	40	22	0.9	325791	11	US-09-768-185A-1	Sequence 1, Appli
	41	21	0.9	21	15	US-10-099-663-3	Sequence 3, Appli
	42	21	0.9	21	15	US-10-099-663-4	Sequence 4, Appli
	43	21	0.9	21	15	US-10-099-663-16	Sequence 16, Appl
	44	21	0.9	133	18	US-10-674-124A-14582	Sequence 14582, A
	45	21	0.9	196	9	US-09-960-352-11421	Sequence 11421, A

## ALIGNMENTS

```

RESULT 1
US-10-099-663-1
; Sequence 1, Application US/10099663
; Publication No. US20030177516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian GUT-Specific Promoters
; FILE REFERENCE: Al81
; CURRENT APPLICATION NUMBER: US/10/099,663
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2381
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(1626)
; NAME/KEY: exon
; LOCATION: (1627)..(1693)
; NAME/KEY: Intron
; LOCATION: (1694)..(2322)
; NAME/KEY: exon
; LOCATION: (2333)..(2381)
US-10-099-663-1

```

Query Match	100.0.%;	Score 2381;	DB 15;	Length 2381;
Best Local Similarity	100.0.%;	Pred. No. 0;		
Matches 2381;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0
Qy	1	AGCTTCCTGCGCAGAAAAGGCTGTGGGGTCTCTCTTCCTCACACAGCTTAAGCAAAATC	60	
Db	1	AGCTTCCTGCGCAGAAAAGGCTGTGGGGTCTCTCTTCCTCACACAGCTTAAGCAAAATC	60	

QY 61 CCCAAGTTCAAAACGTCGGCTGTAAAGGAGATGGCTCACTTCAAATGAAGTGAATATG 120  
Db 61 CCCAAGTTCAAAACGTCGGCTGTAAAGGAGATGGCTCACTTCAAATGAAGTGAATATG 120  
QY 121 AATAATATCAATAAACAGAGCTCTGTGGCAGATCAGAGATAAACCTCTGCTGGGACAAAAT 180  
Db 121 AATAATATCAATAAACAGAGCTCTGTGGCAGATCAGAGATAAACCTCTGCTGGGACAAAAT 180  
QY 181 CTTAAAGTGTATAGGTAGAACAGAGAGTGTTCGCACTAAATGGACTAGATTGTACCACA 240  
Db 181 CTTAAAGTGTATAGGTAGAACAGAGAGTGTTCGCACTAAATGGACTAGATTGTACCACA 240  
QY 241 TTTGATCTCTAGGAGACAAAAGGGTCTGGAAAACAAATTAATCTGCTGGCAGTCAGT 300  
Db 241 TTTGATCTCTAGGAGACAAAAGGGTCTGGAAAACAAATTAATCTGCTGGCAGTCAGT 300  
QY 301 AGCAGCCTGTTTTGGGTGCAACTACAGCAACTTTTGTGCAACAATAACAATCTAAGTTG 360  
Db 301 AGCAGCCTGTTTTGGGTGCAACTACAGCAACTTTTGTGCAACAATAACAATCTAAGTTG 360  
QY 361 TTTTCTTTCTCTCTTTAACTTCTGTACAGTCTAAAGGTGAAGAGTAGCTATTGAGT 420  
Db 361 TTTTCTTTCTCTCTTTAACTTCTGTACAGTCTAAAGGTGAAGAGTAGCTATTGAGT 420  
QY 421 TACTTCCCTCTGCATCTCTTAGCCAGATTAGCATTTGATTTCAAAATGAACCTGAGTGA 480  
Db 421 TACTTCCCTCTGCATCTCTTAGCCAGATTAGCATTTGATTTCAAAATGAACCTGAGTGA 480  
QY 481 ATGGAAAGCCACACTATTTTGGTATCACCAGCAAGTTCTAAATTTATAGTTATCTTCA 540  
Db 481 ATGGAAAGCCACACTATTTTGGTATCACCAGCAAGTTCTAAATTTATAGTTATCTTCA 540  
QY 541 GTAAACCTTTTCTGCGAGTCTGGAAGAAAAGAGATATGATTAACCCAGACTAGTA 600  
Db 541 GTAAACCTTTTCTGCGAGTCTGGAAGAAAAGAGATATGATTAACCCAGACTAGTA 600  
QY 601 AAATTCATTAGTTAGAGCCAAACCTGTATCTGTGTGAAGCAACATTCATTTACAGAT 660  
Db 601 AAATTCATTAGTTAGAGCCAAACCTGTATCTGTGTGAAGCAACATTCATTTACAGAT 660  
QY 661 TCAGGATTTACATTTTGAAGCTAATAGACAGAGATTTGGTCCGCTCCATAGGAACAGA 720  
Db 661 TCAGGATTTACATTTTGAAGCTAATAGACAGAGATTTGGTCCGCTCCATAGGAACAGA 720  
QY 721 CTAACATATATCTGAGTTTAGTACAGCAGATTTAGCACCAAGATTTGCTCAGTTTC 780  
Db 721 CTAACATATATCTGAGTTTAGTACAGCAGATTTAGCACCAAGATTTGCTCAGTTTC 780  
QY 781 AAGTAGCATTATCTTGTGGGAAAGAGAGCTGAGCCAGTGTGTCTCATTTTCTGCATT 840  
Db 781 AAGTAGCATTATCTTGTGGGAAAGAGAGCTGAGCCAGTGTGTCTCATTTTCTGCATT 840  
QY 841 ATCCCTCAACATTTAAACCTGGGATCTATGGAATCAAAACAGTTTGGGTAAATTCAT 900  
Db 841 ATCCCTCAACATTTAAACCTGGGATCTATGGAATCAAAACAGTTTGGGTAAATTCAT 900  
QY 901 TAGCAGCAGATCACTACTGTAGGAATGGACAGAAACAGAGCATTCAGTGAATGGCTAT 960  
Db 901 TAGCAGCAGATCACTACTGTAGGAATGGACAGAAACAGAGCATTCAGTGAATGGCTAT 960  
QY 961 AATATAGAGAAATACGTAGAGGTGTCTGAAATTTAGACTACCTATTAAAGAGTGAGGACA 1020  
Db 961 AATATAGAGAAATACGTAGAGGTGTCTGAAATTTAGACTACCTATTAAAGAGTGAGGACA 1020  
QY 1021 CGAATGGAGAAATATCATTCGAAATTTCTGTAGCTCAGCATAGACTCGAAGGTTTCTGAAA 1080  
Db 1021 CGAATGGAGAAATATCATTCGAAATTTCTGTAGCTCAGCATAGACTCGAAGGTTTCTGAAA 1080  
QY 1081 CTGAACCGAGTTTCCCAAACTACTCTGTGGATGTTTCACTGATCTCTCATCTCATGCTTAT 1140  
Db 1081 CTGAACCGAGTTTCCCAAACTACTCTGTGGATGTTTCACTGATCTCTCATCTCATGCTTAT 1140  
QY 1141 TATGTGGAGTAGAATAGATTCTCACCAAATTAGAATGGACAAACAGAGATTTGTGTTTT 1200

Db 1141 TATGTGGAGTAGAATAGATTCTCACCAAATTAGAATGGACAAACAGAGATTTGTGTTTT 1200  
QY 1201 ATCTGTTGGGTAAATACGTTTTCTCAGTTGTATAAAGCCCTCCACAGTATAAAGTC 1260  
Db 1201 ATCTGTTGGGTAAATACGTTTTCTCAGTTGTATAAAGCCCTCCACAGTATAAAGTC 1260  
QY 1261 CTATGCAACAAGAAATGTCAATACATTTCTTTAGTCTCATTTATTTTTCATTAGATA 1320  
Db 1261 CTATGCAACAAGAAATGTCAATACATTTCTTTAGTCTCATTTATTTTTCATTAGATA 1320  
QY 1321 GCCGGTTTTTACTCAACTCAAAATGAATGAACAGAAATGAATGGGTAGTACTGTTTA 1380  
Db 1321 GCCGGTTTTTACTCAACTCAAAATGAATGAACAGAAATGAATGGGTAGTACTGTTTA 1380  
QY 1381 TAAAGAAGATTAATAAGATCTATCATCTTTAGGCAATTAAGGGAGGAGATTCAG 1440  
Db 1381 TAAAGAAGATTAATAAGATCTATCATCTTTAGGCAATTAAGGGAGGAGATTCAG 1440  
QY 1441 CAAAAGTGTGCTTACAAGTGGAAAACAAGTTTAAATTAAGTGACCCCTCTCTTGACAA 1500  
Db 1441 CAAAAGTGTGCTTACAAGTGGAAAACAAGTTTAAATTAAGTGACCCCTCTCTTGACAA 1500  
QY 1501 GATCAATGCCAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTCCTTCTGATA 1560  
Db 1501 GATCAATGCCAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTCCTTCTGATA 1560  
QY 1561 AGCCTGTTCAATAATTTCTTTGAAAAGCTGTGCTATTCACAGAGTCTGCTACAGAC 1620  
Db 1561 AGCCTGTTCAATAATTTCTTTGAAAAGCTGTGCTATTCACAGAGTCTGCTACAGAC 1620  
QY 1621 AGAAGATGGCAATTTAAACGTTACTTTGAAAATAGAGAAAATGAGAACTATGAAAATTC 1680  
Db 1621 AGAAGATGGCAATTTAAACGTTACTTTGAAAATAGAGAAAATGAGAACTATGAAAATTC 1680  
QY 1681 ATGGAAGCAATGGTGAAGCTTTACTTTTGAATGCCTTCTAAAAAGCAGGATACCACTAC 1740  
Db 1681 ATGGAAGCAATGGTGAAGCTTTACTTTTGAATGCCTTCTAAAAAGCAGGATACCACTAC 1740  
QY 1741 GCGGGAATACAACTTTAAGCTGTTTATGAACTACCATCTGGCTAACCTGCTCTTTGTTGT 1800  
Db 1741 GCGGGAATACAACTTTAAGCTGTTTATGAACTACCATCTGGCTAACCTGCTCTTTGTTGT 1800  
QY 1801 CTGCTATTTTGGCTTGCACATTTGCCCTGCTATTTTGAAGAGACTCTATAGAGGGGA 1860  
Db 1801 CTGCTATTTTGGCTTGCACATTTGCCCTGCTATTTTGAAGAGACTCTATAGAGGGGA 1860  
QY 1861 ATACAAGGAAGAAAACATTTCTGATTTTATTTGCAATGGGATAATCTTATGCAATTTAGCT 1920  
Db 1861 ATACAAGGAAGAAAACATTTCTGATTTTATTTGCAATGGGATAATCTTATGCAATTTAGCT 1920  
QY 1921 AATTCAGTAGAGGATTTCCAGCAGAAATTTAAATAGAAATATATGTAAGGAATATTAT 1980  
Db 1921 AATTCAGTAGAGGATTTCCAGCAGAAATTTAAATAGAAATATATGTAAGGAATATTAT 1980  
QY 1981 TTGATAAGACTCTTTTGAAGAAATTTACAGAGGGGAAATGCTGCTCCAGTTTTCGAGA 2040  
Db 1981 TTGATAAGACTCTTTTGAAGAAATTTACAGAGGGGAAATGCTGCTCCAGTTTTCGAGA 2040  
QY 2041 ACACATGATTTGAGTCAATTTTAACTAGTCTAGTCTTACTTTAAGCTTTGTACAACTGC 2100  
Db 2041 ACACATGATTTGAGTCAATTTTAACTAGTCTAGTCTTACTTTAAGCTTTGTACAACTGC 2100  
QY 2101 CTGTAATATGATGTAACATACTATCTAGTTGGATAGTATTGTAATACAGCTCA 2160  
Db 2101 CTGTAATATGATGTAACATACTATCTAGTTGGATAGTATTGTAATACAGCTCA 2160  
QY 2161 ACACCTGCTCAGTGAAGGGTGGAGAGAGTAAAGACTCTCAGTCAGAAATTCGGGCTAAGC 2220  
Db 2161 ACACCTGCTCAGTGAAGGGTGGAGAGAGTAAAGACTCTCAGTCAGAAATTCGGGCTAAGC 2220  
QY 2221 TCCCTCAACTACAGAAAAGTCACTAATAAATAAGCAATGATGTTCTATTGTTTTT 2280

Db	2221	TCCTCAACTACAGAAAAAGTCACAATAAAAAATGCAACAATGATGTTCTATTTTGTGTTTT	2280
Qy	2281	CTCTGCTTGATGTTAAATGATTATTAATTTTTTTTTTTTAGCGTGAATGTGATGAAAA	2340
Db	2281	CTCTGCTTGATGTTAAATGATTATTAATTTTTTTTTTTTAGCGTGAATGTGATGAAAA	2340
Qy	2341	GAAGTTAGGAGCCAGGATAAATCTGAAGTCTCACTATTTCAG	2381
Db	2341	GAAGTTAGGAGCCAGGATAAATCTGAAGTCTCACTATTTCAG	2381
RESULT 2			
US-10-099-663-2			
; Sequence 2, Application US/10099663			
; Publication No. US20030177516A1			
; GENERAL INFORMATION:			
; APPLICANT: Avigenics, Inc			
; TITLE OF INVENTION: Avian GUT-Specific Promoters			
; FILE REFERENCE: Al81			
; CURRENT APPLICATION NUMBER: US/10/099,663			
; CURRENT FILING DATE: 2002-03-14			
; NUMBER OF SEQ ID NOS: 19			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 2			
; LENGTH: 336			
; TYPE: DNA			
; ORGANISM: Gallus gallus			
; FEATURE:			
; NAME/KEY: misc feature			
; LOCATION: (1)..(336)			
; OTHER INFORMATION: chick intestinal fatty acid binding protein promoter region			
US-10-099-663-2			
Query Match 14.1%; Score 336; DB 15; Length 336;			
Best Local Similarity 100.0%; Pred. No. 4.9e-159;			
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1301	ATTATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAAATAAGATGAACAGAATG	1360
Db	1	ATTATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAAATAAGATGAACAGAATG	60
Qy	1361	AATGGGTAGTGACTGTTTTATAAGAAGAGTAATAAGATACTATCATCTTTTGAGGCAA	1420
Db	61	AATGGGTAGTGACTGTTTTATAAGAAGAGTAATAAGATACTATCATCTTTTGAGGCAA	120
Qy	1421	TAAGGGAGGAGAGATTTCAGCAACACGTGCTTTACAAGTGGAAAAACAAGTTAAACTAAA	1480
Db	121	TAAGGGAGGAGAGATTTCAGCAACACGTGCTTTACAAGTGGAAAAACAAGTTAAACTAAA	180
Qy	1481	GTGACCCCCCTCTCTTGACAAGATCAATGCCACAGTTTGAGCTTTAGCCAGCCACATCATCA	1540
Db	181	GTGACCCCCCTCTCTTGACAAGATCAATGCCACAGTTTGAGCTTTAGCCAGCCACATCATCA	240
Qy	1541	TGTAATTTGCTTTCCTGATGAAGCCTGTTCAATAAATCTCTTTGGAAGCTCTGCTACTTA	1600
Db	241	TGTAATTTGCTTTCCTGATGAAGCCTGTTCAATAAATCTCTTTGGAAGCTCTGCTACTTA	300
Qy	1601	CCAGAAGTCTGCCTACAGACAGAAGAAGATGGCATTTA	1636
Db	301	CCAGAAGTCTGCCTACAGACAGAAGAAGATGGCATTTA	336
RESULT 3			
US-10-221-613-374			
; Sequence 374, Application US/10221613			
; Publication No. US20040029123A1			
; GENERAL INFORMATION:			
; APPLICANT: OLEK, Alexander			
; APPLICANT: PIEPENBROCK, Christian			
; APPLICANT: BERLIN, Kurt			
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle			
; FILE REFERENCE: 5013.1004			
; CURRENT APPLICATION NUMBER: US/10/221,613			





; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Persing, David H.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.566  
; CURRENT APPLICATION NUMBER: US/10/060,036  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 4560  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1770  
; LENGTH: 503  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 373\_374, 480, 481  
; OTHER INFORMATION: n = A,T,C or G  
US-10-060-036-1770

Query Match 1.0%; Score 23; DB 14; Length 503;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 ATTATTATTATTTTTTTTTTTAG 2322  
Db 63 ATTATTATTATTTTTTTTTTTAG 85

## RESULT 10

US-10-311-455-885  
; Sequence 885, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; TITLE OF INVENTION: cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 885  
; LENGTH: 6621  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-885

Query Match 1.0%; Score 23; DB 15; Length 6621;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2297 TTGATTATTATTTTTTTTTT 2319  
Db 5025 TTGATTATTATTTTTTTTTT 5047

## RESULT 11

US-09-764-846-335/c  
; Sequence 335, Application US/09764846

; Patent No. US20020102638A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PTZ12  
; CURRENT APPLICATION NUMBER: US/09/764,846  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 335  
; LENGTH: 10139  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-846-335

Query Match 1.0%; Score 23; DB 9; Length 10139;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2297 TTGATTATTATTTTTTTTTT 2319  
Db 4045 TTGATTATTATTTTTTTTTT 4023

## RESULT 12

US-10-091-483-335/c  
; Sequence 335, Application US/10091483  
; Publication No. US20030049650A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PTZ12C1  
; CURRENT APPLICATION NUMBER: US/10/091,483  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 348  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 335  
; LENGTH: 10139  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-091-483-335

Query Match 1.0%; Score 23; DB 14; Length 10139;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2297 TTGATTATTATTTTTTTTTT 2319  
Db 4045 TTGATTATTATTTTTTTTTT 4023

## RESULT 13

US-10-322-281-706/c  
; Sequence 706, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 706  
; LENGTH: 107543  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-322-281-706

Query Match 1.0%; Score 23; DB 17; Length 107543;



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: January 14, 2005, 18:57:39 ; Search time 7390 Seconds  
(without alignments)  
11740.596 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agctctgcgcgaaagg.....atctgaagctcactattcag 2381

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 18219865908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	586	24.6	1473	8 CC250935	CH261-17C
2	99	4.2	598	6 CF250818	esa017_e0
3	95	4.0	595	6 CF250947	esa019_a0
4	85	3.6	705	5 BU35106	603474288
5	84	3.5	759	5 BU265304	603508640
6	81	3.4	829	5 BU296697	603741965
7	79	3.3	814	5 BU234099	603792609
8	79	3.3	885	5 BU123280	603003274
9	78	3.3	1080	5 BU123336	603149475
10	74	3.1	576	6 CD739141	4026481_1
11	62	2.6	426	6 CD739695	4028031_1
12	60	2.5	443	5 BX276255	BK276255
13	60	2.5	480	6 CD733346	4045132_1
14	60	2.5	885	5 BU123717	603147391
c 15	26	1.1	673	9 AG096827	Pan trogl
16	25	1.0	470	8 AQ942982	Sheared D
17	25	1.0	761	8 BZ514427	BOMSG62TF
18	24	1.0	118	5 BP741179	BP741179
19	24	1.0	334	1 AA112689	BP741179
20	24	1.0	485	2 BF024733	dc86e08.x
c 21	24	1.0	520	1 AL598814	DKFZp313J
22	24	1.0	522	4 BG161028	dab64d06
c 23	24	1.0	639	9 TA120F06Q	AL462904 T. brucei
c 24	24	1.0	643	2 BB646067	BB646067

ALIGNMENTS

RESULT 1  
CC250935/c  
LOCUS CH261-17C7\_Sp6.1 CH261 Gallus gallus genomic clone CH261-17C7, 1473 bp DNA linear GSS 13-MAY-2003  
DEFINITION genomic survey sequence.  
ACCESSION CC250935  
VERSION CC250935.1 GI:30587685  
KEYWORDS GSS.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 1473)  
AUTHORS Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.  
TITLE Gallus gallus BAC End Reads  
JOURNAL Unpublished (2003)  
COMMENT Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: submissions@wustl.edu  
Insert Length: 182000 Std Error: 0.00  
Seq primer: Sp6 ATTTAGTGACACTATAG  
Class: BAC ends  
High quality sequence start: 239  
High quality sequence stop: 912.  
Location/Qualifiers  
source  
1. .1473  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/strain="Red Jungle Fowl"  
/db\_xref="taxon:9031"  
/clone="CH261-17C7"  
/sex="female"  
/cell\_line="UCD001, inbred 256"  
/clone\_lib="CH261"  
/notes="Vector: pFARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: http://www.chori.org/bacpac"

ORIGIN

Query Match 24.8%; Score 586; DB 8; Length 1473;  
Best Local Similarity 99.8%; Pred. No. 4e-293;  
Matches 636; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1684 GAAGCAATGGTAAAGCCTTACTTTTGAATGCTTCTTAAAGCAGGATACCTACGCG 1743

Db 936 GAAGCAATGGTAAAGCCTTACTTTTGAATGCTCTTAAAGCAGATACCACTACGGC 877  
QY 1744 GGAATACAAACTTAAGCTGTTCATGAACTACCATCTGGCTAACCTGCTCTTTGTTGTCG 1803  
Db 876 GGAATACAAACTTAAGCTGTTCATGAACTACCACTGGCTAACCTGCTCTTTGTTGTCG 817  
QY 1804 CTATTTTGGCCCTTGCACATTCGCCCTGCACCTATTATTTGAAAAGACTCTATAGAGGGGAATA 1863  
Db 816 CTATTTTGGCCCTTGCACATTCGCCCTGCACCTATTATTTGAAAAGACTCTATAGAGGGGAATA 757  
QY 1864 CAAGGAAGAAAAACATTCCTGATTTTATTTGCAATGGGATAATCTTATGCAATTTAGCTAAT 1923  
Db 756 CAAGGAAGAAAAACATTCCTGATTTTATTTGCAATGGGATAATCTTATGCAATTTAGCTAAT 697  
QY 1924 TCAGTAGAGGCAATTCAGCAGAAAAATTAATAGAAATATATGTAAGGAATATATTATTG 1983  
Db 696 TCAGTAGAGGCAATTCAGCAGAAAAATTAATAGAAATATATGTAAGGAATATATTATTG 637  
QY 1984 ATAAGACTGTTTGAAGAAATACAGAGGGAATTCGCTGCTCCAGTTTTCGAGAAC 2043  
Db 636 ATAAGACTGTTTGAAGAAATACAGAGGGAATTCGCTGCTCCAGTTTTCGAGAAC 577  
QY 2044 CACATGATTTGAGCTCAATTTTAAACATGCTAGTCTTACTTTAAGCTTTGTACAACTGCCTG 2103  
Db 576 CACATGATTTGAGCTCAATTTTAAACATGCTAGTCTTACTTTAAGCTTTGTACAACTGCCTG 517  
QY 2104 TAATATGGATGTAACAACTATCTAGTGTGATAGTAGTTGTAATACAGCTGAACA 2163  
Db 516 TAATATGGATGTAACAACTATCTAGTGTGATAGTAGTTGTAATACAGCTGAACA 457  
QY 2164 CTGCTCAGTGAAGGTGGAGAGAGTAAAGCTCTGAGTCAGATCTGGGCTAAAGCTCC 2223  
Db 456 CTGCTCAGTGAAGGTGGAGAGAGTAAAGCTCTGAGTCAGATCTGGGCTAAAGCTCC 397  
QY 2224 CTCAACTACAGAAAAAGTCAAAATGCAAAATGCAAAATGATGTTCTATTGTTTTC 2283  
Db 396 CTCAACTACAGAAAAAGTCAAAATGCAAAATGCAAAATGATGTTCTATTGTTTTC 337  
QY 2284 TGCTTGATGTAATGATTAATTAATTTTATTTT 2320  
Db 336 TGCTTGATGTAATGATTAATTAATTTTATTTT 300

RESULT 2  
CF250818 598 bp mRNA linear EST 07-AUG-2003  
LOCUS esa017\_e05 Eimeiria tenella-infected caecal tonsil Gallus gallus  
DEFINITION CDNA, mRNA sequence.  
ACCESSION CF250818  
VERSION CF250818.1 GI:33484073  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Archosauaria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
TITLE 1 (bases 1 to 598)  
JOURNAL Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,  
COMMENT A collection of chicken ESTs from activated immune cells  
Unpublished (2003)  
Contact: Zoorob R  
UPR 1983  
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@vjf.cnrs.fr.  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"

ORIGIN  
Query Match 4.2%; Score 99; DB 6; Length 598;  
Best Local Similarity 100.0%; Pred. No. 2e-39;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1596 ACTTACCAGAACTCTGCCTACAGACAGAAAGATGGCAATTTAACGGTACTTGGAAATAGA 1655  
Db 3 ACTTACCAGAACTCTGCCTACAGACAGAAAGATGGCAATTTAACGGTACTTGGAAATAGA 62  
QY 1656 GAAATATGAGAACTATGAAAAATTCATGGAAGCAATGGG 1694  
Db 63 GAAATATGAGAACTATGAAAAATTCATGGAAGCAATGGG 101

RESULT 3  
CF250947 595 bp mRNA linear EST 07-AUG-2003  
LOCUS esa019\_a07 Eimeiria tenella-infected caecal tonsil Gallus gallus  
DEFINITION CDNA, mRNA sequence.  
ACCESSION CF250947  
VERSION CF250947.1 GI:33484202  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Archosauaria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
TITLE 1 (bases 1 to 595)  
JOURNAL Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,  
COMMENT A collection of chicken ESTs from activated immune cells  
Unpublished (2003)  
Contact: Zoorob R  
UPR 1983  
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@vjf.cnrs.fr.  
Location/Qualifiers  
1. .595  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone\_lib="Eimeiria tenella-infected caecal tonsil"  
/note="Organ: Caecal tonsil; Vector: pTriplex2"

ORIGIN  
Query Match 4.0%; Score 95; DB 6; Length 595;  
Best Local Similarity 100.0%; Pred. No. 2.4e-37;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1600 ACCAGAACTCTGCCTACAGACAGAAAGATGGCAATTTAACGGTACTTGGAAATAGA 1659  
Db 3 ACCAGAACTCTGCCTACAGACAGAAAGATGGCAATTTAACGGTACTTGGAAATAGA 62  
QY 1660 AATGAGAACTATGAAAAATTCATGGAAGCAATGGG 1694  
Db 63 AATGAGAACTATGAAAAATTCATGGAAGCAATGGG 97

RESULT 4  
BU355106 705 bp mRNA linear EST 28-NOV-2002  
LOCUS 603474288F1 CSEQCHN70 Gallus gallus CDNA clone Chest355022 5', mRNA  
DEFINITION sequence.  
ACCESSION BU355106  
VERSION BU355106.1 GI:25863107  
KEYWORDS EST.

**SOURCE**  
 Gallus gallus (chicken)  
**ORGANISM**  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
**REFERENCE**  
 1 (bases 1 to 705)  
**AUTHORS**  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE**  
 A Comprehensive Collection of Chicken cDNAs  
**JOURNAL**  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE**  
 22335534  
**PUBMED**  
 12445392  
**COMMENT**  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
**FEATURES**  
 source  
 1..705  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST355022"  
 /dev\_stage="36"  
 /lab\_host="DH10B"  
 /clone\_lib="CSQCIN70"  
 /note="Organ: hearts; Vector: pBluescript II KS(+); Site 1:  
 Site 1: EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."  
**ORIGIN**  
 Query Match 3.6%; Score 85; DB 5; Length 705;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-32;  
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1610 TGCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACT 1669  
 Db 1 TGCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACT 60  
 QY 1670 ATGAAAAATTCATGGAAGCAATGGG 1694  
 Db 61 ATGAAAAATTCATGGAAGCAATGGG 85  
**RESULT 5**  
 BU265304 759 bp mRNA linear EST 26-NOV-2002  
**LOCUS**  
 603508640F1 CSEQCHN52 Gallus gallus cDNA clone CHEST437010 5', mRNA  
**DEFINITION**  
 sequence.  
**ACCESSION**  
 BU265304  
**VERSION**  
 BU265304.1 GI:25536254  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Gallus gallus (chicken)  
**ORGANISM**  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
**REFERENCE**  
 1 (bases 1 to 759)  
**AUTHORS**  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE**  
 A Comprehensive Collection of Chicken cDNAs  
**JOURNAL**  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE**  
 22335534  
**PUBMED**  
 12445392  
**COMMENT**  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
**FEATURES**  
 source  
 1..759  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST437010"  
 /dev\_stage="22"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN52"  
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site 1:  
 EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."  
**ORIGIN**  
 Query Match 3.5%; Score 84; DB 5; Length 759;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-31;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1611 GCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTA 1670  
 Db 1 GCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTA 60  
 QY 1671 TGAATAATTCATGGAAGCAATGGG 1694  
 Db 61 TGAATAATTCATGGAAGCAATGGG 84  
**RESULT 6**  
 BU296697 829 bp mRNA linear EST 27-NOV-2002  
**LOCUS**  
 603741965F1 CSEQCHN56 Gallus gallus cDNA clone CHEST640b17 5', mRNA  
**DEFINITION**  
 sequence.  
**ACCESSION**  
 BU296697  
**VERSION**  
 BU296697.1 GI:25746333  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Gallus gallus (chicken)  
**ORGANISM**  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
**REFERENCE**  
 1 (bases 1 to 829)  
**AUTHORS**  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE**  
 A Comprehensive Collection of Chicken cDNAs  
**JOURNAL**  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE**  
 22335534  
**PUBMED**  
 12445392  
**COMMENT**  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
**FEATURES**  
 source  
 1..759  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST437010"  
 /dev\_stage="22"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN52"  
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site 1:  
 EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."  
**ORIGIN**  
 Query Match 3.5%; Score 84; DB 5; Length 759;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-31;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1611 GCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTA 1670  
 Db 1 GCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTA 60  
 QY 1671 TGAATAATTCATGGAAGCAATGGG 1694  
 Db 61 TGAATAATTCATGGAAGCAATGGG 84  
**RESULT 6**  
 BU296697 829 bp mRNA linear EST 27-NOV-2002  
**LOCUS**  
 603741965F1 CSEQCHN56 Gallus gallus cDNA clone CHEST640b17 5', mRNA  
**DEFINITION**  
 sequence.  
**ACCESSION**  
 BU296697  
**VERSION**  
 BU296697.1 GI:25746333  
**KEYWORDS**  
 EST.  
**SOURCE**  
 Gallus gallus (chicken)  
**ORGANISM**  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
**REFERENCE**  
 1 (bases 1 to 829)  
**AUTHORS**  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
**TITLE**  
 A Comprehensive Collection of Chicken cDNAs  
**JOURNAL**  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
**MEDLINE**  
 22335534  
**PUBMED**  
 12445392  
**COMMENT**  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (

## COMMENT

Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers

FEATURES  
source

1..829  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton Line 151"  
 /db\_xref="taxon:9031"  
 /clone="CHEST640D17"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN56"  
 /note="Organ: small intestine; Vector: pBluescript II  
 KS(+); Site\_1: EcoRI; Site\_2: NotI; This normalized  
 library was constructed from 1 million independent clones.  
 cDNA synthesis was initiated using an oligo(dT) primer,  
 using methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 3.4%; Score 81; DB 5; Length 829;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-30;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1614 TACGACAGAAAGATGCGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTATGA 1673  
 Db 11 TACGACAGAAAGATGCGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTATGA 70  
 QY 1674 AAAATTCATGGAAGCAATGGG 1694  
 Db 71 AAAATTCATGGAAGCAATGGG 91

RESULT 7  
 BU234099  
 LOCUS 603792609F1 CSEQCHN24 Gallus gallus cDNA clone CHEST758b4 5', mRNA  
 DEFINITION  
 sequence.  
 ACCSSION  
 BU234099  
 VERSION BU234099.1 GI:25478348  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK

FEATURES  
source

1..814  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST758b4"  
 /dev\_stage="22"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN24"  
 /note="Organ: heads; Vector: pBluescript II KS(+); Site\_1:  
 EcoRI; Site\_2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 3.3%; Score 79; DB 5; Length 814;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-29;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1616 CAGACAGAAAGATGCGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTATGAAA 1675  
 Db 1 CAGACAGAAAGATGCGCATTTAACGGTACTTGGAAAATAGAGAAAATGAGAACTATGAAA 60  
 QY 1676 AATTCATGGAAGCAATGGG 1694  
 Db 61 AATTCATGGAAGCAATGGG 79

RESULT 8  
 BU123280  
 LOCUS 603003274F1 CSEQCHL18 Gallus gallus cDNA clone CHEST14c20 5', mRNA  
 DEFINITION  
 sequence.  
 ACCESSION  
 BU123280  
 VERSION BU123280.1 GI:25333903  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers

FEATURES  
source

1..885  
 /organism="Gallus gallus"

```

/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST14c20"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHL18"
/notes="Organ: small intestine; Vector: pBluescript II
KS(+); Site 1: EcoRI; Site 2: NotI; Modification of
pBluescript_II KS(+) [Stratagene] vector to accommodate
cDNA produced with the T-trimmed protocol (Construction of
uni-directionally cloned cDNA libraries from messenger RNA
for improved 3' end DNA sequencing by Glenn Fu, et al.
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
NotI and EcoRI. Ligate in double stranded adaptor
containing BglI and BamHI sites
[5'ggcgcgctgcagcccgatccgcaaaaaag]
[5'aattcttttttcggtccgggtgcgcg]"

ORIGIN
Query Match 3.3%; Score 79; DB 5; Length 885;
Best Local Similarity 100.0%; Pred. No. 5.2e-29;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CAGACAGAAAGATGCGCATTTAAACGGTACTTGGAAAATAGAGAAAAATGAGAACTATGAAA 1675
Db 3 CAGACAGAAAGATGCGCATTTAAACGGTACTTGGAAAATAGAGAAAAATGAGAACTATGAAA 62

QY 1676 AATTCATGGAAGCAATGGG 1694
Db 63 AATTCATGGAAGCAATGGG 81

RESULT 9
LOCUS BUI23336 1080 bp mRNA linear EST 25-NOV-2002
DEFINITION 603149475F1 CSEQCHL18 Gallus gallus CDNA clone CHEST153f19 5', mRNA
sequence.
BUI23336
VERSION BUI23336.1 GI:253333972
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1080)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE 12445392
PUBMED
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(MUMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..1080
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST153f19"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHL18"

```

```

/notes="Organ: small intestine; Vector: pBluescript II
KS(+); Site 1: EcoRI; Site 2: NotI; Modification of
pBluescript_II KS(+) [Stratagene] vector to accommodate
cDNA produced with the T-trimmed protocol (Construction of
uni-directionally cloned cDNA libraries from messenger RNA
for improved 3' end DNA sequencing by Glenn Fu, et al.
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
NotI and EcoRI. Ligate in double stranded adaptor
containing BglI and BamHI sites
[5'ggcgcgctgcagcccgatccgcaaaaaag]
[5'aattcttttttcggtccgggtgcgcg]"

ORIGIN
Query Match 3.3%; Score 78; DB 5; Length 1080;
Best Local Similarity 100.0%; Pred. No. 1.7e-28;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1617 AGACAGAAAGATGCGCATTTAAACGGTACTTGGAAAATAGAGAAAAATGAGAACTATGAAA 1676
Db 11 AGACAGAAAGATGCGCATTTAAACGGTACTTGGAAAATAGAGAAAAATGAGAACTATGAAA 70

QY 1677 AATTCATGGAAGCAATGGG 1694
Db 71 AATTCATGGAAGCAATGGG 88

RESULT 10
LOCUS CD739141 576 bp mRNA linear EST 26-JUN-2003
DEFINITION 4026481 lGAL - Chicken Intestinal lymphocyte Gallus gallus CDNA
clone lGAL_93A12 5', mRNA sequence.
CD739141
ACCESSION CD739141
VERSION CD739141.1 GI:32289990
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 576)
Min,W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van
Tassel,C. and Han,J.Y.
Chicken intestinal lymphocyte EST database as a resource for the
analysis of mucosal immune function
Unpublished (2003)
JOURNAL
COMMENT Contact: Hyun S. Lillehoj
Animal Parasite Diseases Laboratory
Animal and Natural Resources Institute, USDA
Bldg.1043, BARC-East, Beltsville, MD 20705, USA
Tel: 3015048771
Fax: 3015045103
Email: hlilleho@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt -, -trim fasta. Vector identified
by cross match using options -minmatch 12 -minscore 12
Plate: 93 row: A column: 12
Seq primer: ATTTAGTGCACATATAG
High quality sequence stop: 576.
Location/Qualifiers
1..576
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="white leghorn SC"
/db_xref="taxon:9031"
/clone="lGAL_93A12"
/sex="mixed"
/tissue_type="Gut"
/cell_type="lymphocyte"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="lGAL - Chicken Intestinal Lymphocyte"
/notes="Organ: Intestine; Vector: pCMV-SPORT6; Site 1:
SalI; Site 2: NotI; Normalized library from chicken gut

```

infected with coccidia duodenum and middle gut."

ORIGIN  
Query Match 3.1%; Score 74; DB 6; Length 576;  
Best Local Similarity 100.0%; Pred. No. 2.1e-26;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1621 AGAAATGCGATTTAACGGTACTTGGAAAATAGAAAAATGAGAACTATGAAAAATTC 1680  
|||||  
Db 1 AGAAGATGCGATTTAACGGTACTTGGAAAATAGAAAAATGAGAACTATGAAAAATTC 60  
|||||

QY 1681 ATGGAAGCAATGGG 1694  
|||||  
Db 61 ATGGAAGCAATGGG 74  
|||||

RESULT 11  
CD739695 426 bp mRNA linear EST 26-JUN-2003  
LOCUS 4028031 1GAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA  
DEFINITION clone 1GAL\_20B02 5', mRNA sequence.  
CD739695  
ACCESSION CD739695.1 GI:32290544  
VERSION  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 426)  
TITLE Min.W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van  
Tassel,C. and Han,J.Y.  
JOURNAL Chicken intestinal lymphocyte EST database as a resource for the  
COMMENT analysis of mucosal immune function  
Unpublished (2003)  
Contact: Hyun S. Lillehoj  
Animal Parasite Diseases Laboratory  
Animal and Natural Resources Institute, USDA  
Bldg.1043, BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048771  
Fax: 3015045103  
Email: hlilleho@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim\_alt '-trim fasta. Vector identified  
by cross\_match using options -minmatch 12 -minscore 12  
Plate: 20 row: B column: 02  
Seq primer: ATTAGGTGACACTATAG  
High quality sequence stop: 426.  
FEATURES  
source  
1..426  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="white leghorn SC"  
/db\_xref="taxon:9031"  
/clone="1GAL\_20B02"  
/sex="mixed"  
/tissue\_type="Gut"  
/cell\_type="Lymphocyte"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="1GAL - Chicken Intestinal Lymphocyte"  
/notes="Organ: Intestine; Vector: pCMV-SPORT6; Site:1;  
Sall; Site2: NotI; Normalized library from chicken gut  
infected with coccidia duodenum and middle gut."

ORIGIN  
Query Match 2.6%; Score 62; DB 6; Length 426;  
Best Local Similarity 100.0%; Pred. No. 3.6e-20;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1633 TTTAACGGTACTTGGAAAATAGAAAAATGAGAACTATGAAAAATTCATGGAAGCAATG 1692  
|||||  
Db 19 TTTAACGGTACTTGGAAAATAGAAAAATGAGAACTATGAAAAATTCATGGAAGCAATG 78  
|||||

QY 1693 GG 1694  
||  
Db 79 GG 80  
||

RESULT 12  
BX276255 443 bp mRNA linear EST 15-JUL-2004  
LOCUS BX276255 AGENAE Gallus gallus multi-tissues normalized library  
DEFINITION (gcag) Gallus gallus cDNA clone gcag0008c.c.24 5prim, mRNA  
sequence.  
ACCESSION BX276255  
VERSION BX276255.1 GI:28598746  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 443)  
AUTHORS Herault,F., Le Meuth-Metzinger,V., Desert,C., Retout,B., Piumi,F.,  
Klopp,C. and Douaire,M.  
TITLE Construction and primary characterization of chicken normalized  
JOURNAL multi-tissue cDNA libraries  
COMMENT Unpublished (2003)  
Contact: Douaire M  
INRA, UMR INRA-ENSAR Genetique Animale  
65, rue de Saint-Brieuc, RENNES cedex, 35042, FRANCE  
Tel: +33 (0) 2.23.48.54.63  
Fax: +33 (0) 2.23.48.54.70  
Email: Madeleine.Douaire@roazhon.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us  
at signenaupport@jouy.inra.fr to obtain the chromatogram of this  
sequence.  
Seq primer: M13R.  
Plate: 0008 row: C column: 24  
FEATURES  
source  
1..443  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="gcag0008c.c.24"  
/tissue\_type="multi-tissues"  
/dev\_stage="from embryos to adults"  
/lab\_host="DH10B"  
/clone\_lib="AGENAE Gallus gallus multi-tissues normalized  
library (gcag)"  
/note="Vector: pT7T3D-pac; tissues: brain, embryo,  
kidney, multi-tissues, muscle, pancreas, skin, testis,  
liver, adipose tissue, granulosa, utero-vaginal gland,  
oviduct, small follicle, ovary, hypothalamus, pituitary  
gland, ileon, jejunum, caecum, duodenum, spleen,  
fabricius gland, bone marrow, thymus, hematopoietic  
progenitor cells. Clone distribution : AGENAE Resource  
centre. Francois PIUMI, Francois.Piumi.inra.fr, INRA, CEA  
Radiobiologie et Etude du genome (LREG), Domaine de  
Vilvert, 78352, Jouy-en-Josas cedex, FRANCE"

Query Match 2.5%; Score 60; DB 5; Length 443;  
Best Local Similarity 100.0%; Pred. No. 4e-19;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2322 GCGGTGAATGTGATGAAAAGAAAGTTAGAGCCCGATATCTGAAGTCACTATTTCAG 2381  
|||||  
Db 33 GCGGTGAATGTGATGAAAAGAAAGTTAGAGCCCGATATCTGAAGTCACTATTTCAG 92  
|||||

RESULT 13  
CD733346 480 bp mRNA linear EST 26-JUN-2003  
LOCUS CD733346  
DEFINITION 4045132 1GAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA



clone 1GAL\_41E15 5', mRNA sequence.

ACCESSION CD733346  
 VERSION CD733346.1 GI:32284195  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 480)  
 AUTHORS Min, W., Lillehoj, H. S., Ashwell, C. M., Matukumalli, L. K., van Tassel, C. and Han, J. Y.  
 TITLE Chicken intestinal lymphocyte EST database as a resource for the analysis of mucosal immune function  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Hyun S. Lillehoj  
 Animal Parasite Diseases Laboratory  
 Animal and Natural Resources Institute, USDA  
 Bldg. 1043, BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048771  
 Fax: 3015045103  
 Email: lillehoj@anri.barc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred 0.000925 using options -trim alt -, -trim fasta. Vector identified by cross match using options -minmatch 12 -minscore 12  
 Plate: 41 row: E column: 15  
 Seq primer: ATTAGGTGACACTATAG  
 High quality sequence stop: 480.  
 Location/Qualifiers  
 1. .480  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="white leghorn SC"  
 /db\_xref="taxon:9031"  
 /clone="1GAL\_41E15"  
 /sex="mixed"  
 /tissue\_type="Gut"  
 /cell\_type="Lymphocyte"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="1GAL - Chicken Intestinal Lymphocyte"  
 /note="Organ: Intestine; Vector: pCMV-SPORT6; Site 1: Salt; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."

ORIGIN  
 Query Match 2.5%; Score 60; DB 6; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 4e-19;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2322 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 2381  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 71 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 130

RESULT 14  
 BUI23717  
 LOCUS BUI23717 885 bp mRNA linear EST 25-NOV-2002  
 DEFINITION 603147391P1 CSEQCHL18 Gallus gallus CDNA clone CHEST148f24 5', mRNA sequence.  
 ACCESSION BUI23717  
 VERSION BUI23717.1 GI:25334366  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 885)  
 AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)

clone 1GAL\_41E15 5', mRNA sequence.

ACCESSION CD733346  
 VERSION CD733346.1 GI:32284195  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 480)  
 AUTHORS Min, W., Lillehoj, H. S., Ashwell, C. M., Matukumalli, L. K., van Tassel, C. and Han, J. Y.  
 TITLE Chicken intestinal lymphocyte EST database as a resource for the analysis of mucosal immune function  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Hyun S. Lillehoj  
 Animal Parasite Diseases Laboratory  
 Animal and Natural Resources Institute, USDA  
 Bldg. 1043, BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048771  
 Fax: 3015045103  
 Email: lillehoj@anri.barc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred 0.000925 using options -trim alt -, -trim fasta. Vector identified by cross match using options -minmatch 12 -minscore 12  
 Plate: 41 row: E column: 15  
 Seq primer: ATTAGGTGACACTATAG  
 High quality sequence stop: 480.  
 Location/Qualifiers  
 1. .480  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="white leghorn SC"  
 /db\_xref="taxon:9031"  
 /clone="1GAL\_41E15"  
 /sex="mixed"  
 /tissue\_type="Gut"  
 /cell\_type="Lymphocyte"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="1GAL - Chicken Intestinal Lymphocyte"  
 /note="Organ: Intestine; Vector: pCMV-SPORT6; Site 1: Salt; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."

ORIGIN  
 Query Match 2.5%; Score 60; DB 6; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 4e-19;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2322 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 2381  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 71 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 130

RESULT 14  
 BUI23717  
 LOCUS BUI23717 885 bp mRNA linear EST 25-NOV-2002  
 DEFINITION 603147391P1 CSEQCHL18 Gallus gallus CDNA clone CHEST148f24 5', mRNA sequence.  
 ACCESSION BUI23717  
 VERSION BUI23717.1 GI:25334366  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 885)  
 AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)

22335534  
 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers  
 1. .885  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton Line 151"  
 /db\_xref="taxon:9031"  
 /clone="CHEST148f24"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHL18"  
 /note="Organ: small intestine; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+) [Stratagene] vector to accommodate cDNA produced with the T-trimmed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspI and BamHI sites  
 [5'ggcgcgtgcagccgcggtccgaaacaaag]  
 [5'aattcttttttcggatccggcggtgcgcg]"

ORIGIN  
 Query Match 2.5%; Score 60; DB 5; Length 885;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-19;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2322 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 2381  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 81 GGCGTGAATGTGATGAAGAAGTATAGGAGCCACGATATCTGAAGCTCATTTCAG 140

RESULT 15  
 AG096827/c  
 LOCUS AG096827 673 bp DNA linear GSS 03-NOV-2001  
 DEFINITION Pan troglodytes DNA, clone: PTB-098F02.F, genomic survey sequence.  
 ACCESSION AG096827  
 VERSION AG096827.1 GI:16717344  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes

REFERENCE 1  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of Library PTB  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 673)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-Chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
 PRIMERS  
 Sequencing: -21M13

## LIBRARY

Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.

## FEATURES

```
Location/Qualifiers
1. .673
/organisms="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-098F02.F"
/sex="male"
/cell_type="lymphoblast"
/cell_line="PTB Chimpanzee M
```

## ORIGIN

Query Match 1.1%; Score 26; DB 9; Length 673;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2296 ATTGATTATTA TTTTITTTA 2321  
Db 190 ATTGATTATTA TTTTITTTA 165

Search completed: January 15, 2005, 00:43:44  
Job time : 7399 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 07:35:21 ; Search time 10545.8 Seconds  
(without alignments)  
10676.925 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agttctgcgcagaaag.....atctgaagtcactattcag 2381

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues  
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	736.8	30.9	2847	5	AY254202	AY254202 Gallus ga
2	85	3.6	923	5	BX932340	BX932340 Gallus ga
3	81	3.4	899	5	BX930518	BX930518 Gallus ga
4	78.8	3.3	896	5	BX933597	BX933597 Gallus ga
5	64.2	2.7	1141	6	AX083744	AX083744 Sequence
6	62	2.6	3013	10	RATFABP	M18080 Rat intesti
7	62	2.6	230897	2	AC130496	AC130496 Rattus no
8	57.8	2.4	1141	6	AX083744	AX083744 Sequence
9	56.6	2.4	5039	10	MUSGFABP	M65033 Mouse Fabp
10	51.8	2.2	234081	3	PFMAL1P2	AL035475 Plasmodiu
11	50	2.1	564	10	RATFABP	M35992 Rat intesti
12	49.8	2.1	619	5	XELIFABP	L19946 Xenopus lae
13	49.8	2.1	5204	9	HUMFABP	AX771590 Sequence
14	49.8	2.1	5204	9	HUMFABP	M18079 Human, inte
15	49.8	2.1	200000	2	AC008077	AC008077 Homo sapi
16	49.2	2.1	2000	6	AX655393	AX655393 Sequence
17	49	2.1	136061	2	AC117821	AC117821 Mus muscu
18	49	2.1	202804	10	AC021630	AC021630 Mus muscu
19	49	2.1	249943	3	AE014823	AE014823 Plasmodiu

20	48.4	2.0	564	10	RATFABP	K01180 Rat intesti
C 21	48.4	2.0	832	6	AR415176	AR415176 Sequence
C 22	48.4	2.0	832	6	AX972010	AX972010 Sequence
C 23	48.4	2.0	832	6	BD110729	BD110729 EST and e
C 24	48.4	2.0	18876	3	CEF32G8	Z72509 Caenorhabdi
C 25	48.4	2.0	197110	9	AC104306	AC104306 Homo sapi
C 26	48.2	2.0	132254	3	AC116330	AC116330 Dictyoste
C 27	48.2	2.0	132254	3	AC116330	AC116330 Dictyoste
C 28	48	2.0	191840	9	AC092656	AC092656 Homo sapi
C 29	47.8	2.0	110000	3	AC116984_1	Continuation (2 of
C 30	47.8	2.0	137174	2	BX276186	BX276186 Danio rer
C 31	47.8	2.0	162810	2	CR392028	CR392028 Danio rer
C 32	47.6	2.0	5302	3	AC116961	AC116961 Dictyoste
C 33	47.6	2.0	143364	2	BX324220	BX324220 Danio rer
C 34	47.6	2.0	174176	5	AL772340	AL772340 Zebrafish
C 35	47.6	2.0	177623	2	CR394571	CR394571 Danio rer
C 36	47.4	2.0	110000	2	PFMAL13_09	Continuation (10 o
C 37	47.4	2.0	318221	2	PFMAL13P3	AL049184 Plasmodiu
C 38	47	2.0	2636	6	CQ593303	CQ593303 Sequence
C 39	47	2.0	81181	2	AC019668	AC019668 Drosophil
C 40	47	2.0	176969	3	AC010707	AC010707 Drosophil
C 41	47	2.0	181815	5	BX248120	BX248120 Zebrafish
C 42	47	2.0	184977	5	BX649334	BX649334 Zebrafish
C 43	47	2.0	192540	3	AC010846	AC010846 Drosophil
C 44	47	2.0	220177	2	CR354597	CR354597 Danio rer
C 45	47	2.0	323844	3	AE003501	AE003501 Drosophil

ALIGNMENTS

RESULT 1	AY254202	2847 bp	DNA	linear	VRT 22-APR-2003
LOCUS	Gallus gallus intestinal fatty acid-binding protein gene, complete				
DEFINITION	cds.				
ACCESSION	AY254202				
VERSION	AY254202.1	GI:30060211			
KEYWORDS	Gallus gallus (chicken)				
SOURCE	Gallus gallus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (bases 1 to 2847)				
AUTHORS	Wang,Q., Li,H., Wang,Y. and Zhao,J.				
TITLE	Cloning and characterization of chicken I-FABP gene				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2847)				
AUTHORS	Wang,Q., Li,H., Wang,Y. and Zhao,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-MAR-2003) Animal Science & Technology, Northeast Agricultural University, 59 Mucai Street, Harbin, Heilongjiang 150030, China				

FEATURES	Location/Qualifiers
source	1..2847
	/organism="Gallus gallus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9031"
	/tissue_type="blood"
	/notes="Breed: Arber Acres broiler"
mRNA	join(<21..87,720..892,1465..1572,2416..>2466)
	/product="intestinal fatty acid-binding protein"
CDS	join(21..87,720..892,1465..1572,2416..2466)
	/codon_start=1
	/product="intestinal fatty acid-binding protein"
	/protein_id="AAP13101.1"
	/db_xref="GI:30060212"
	/translation="MAPNGTWKIEKNYKFWKANGVNVMMKKLGAHDNKLKTIQDD
	GNKFLKSSNFRIDIEFTLGVFSFYSLADGTELSGNWLNKLVGTFTRKDNKGV
	L7AYREIVGSELITQTVYVEGVEAKRIFKE"
ORIGIN	

```
Query Match      30.9%; Score 736.8; DB 5; Length 2847;
Best Local Similarity 98.1%; Pred. No. 4e-146;
Matches 757; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

Qy 1613 CTACAGACAGAAAGATGGCAATTTAAACGGTACTTGGAAATATAGAGAAATATGAGAACTATG 1672
Db 7 CGAGAGACAGAAAGATGGCAATTTAAACGGTACTTGGAAATATAGAGAAATATGAGAACTATG 66

Qy 1673 AAAAAATTCATGGAAGCAATGGTAAAGCCCTTACTTTTTTGAATGCTTCTAAAAAGCAGGAT 1732
Db 67 AAAAAATTCATGGAAGCAATGGTAAAGCCCTTACTTTTTTGAATGCTTCTAAAAAGCAGGAT 126

Qy 1733 ACCACTACGGGGGATACAACTTAAGCTGTTCAAGAACTACCACTACGCTACGCTACCTGTC 1792
Db 127 ACCACTACGGGGGATACAACTTAAGCTGTTCAAGAACTACCACTACGCTACGCTACCTGTC 186

Qy 1793 TTTGTTGCTGCTATTTTGGCCCTTGCACTTGCCCTGCCTTATTTTGAAGAACTCTAT 1852
Db 187 TTTGTTGCTGCTATTTTGGCCCTTGCACTTGCCCTGCCTTATTTTGAAGAACTCTAT 246

Qy 1853 AGGGGGAATACAAAGGAAGAAAACATCTGATTTTATTTGCAATGCGATAATCTTATGC 1912
Db 247 AGGGGGAATACAAAGGAAGAAAACATCTGATTTTATTTGCAATGCGATAATCTTATGC 306

Qy 1913 ATTTAGCTAATTCAGTAGAGGATCCAGCAGAAATTTAAATAGAAATATATATGAAGA 1972
Db 307 ATTTAGCTAATTCAGTAGAGGATCCAGCAGAAATTTAAATAGAAATATATATGAAGA 366

Qy 1973 ATATTATTTTGATAAGACTGTTTGAAATATACACAGAGGAAATTCGTGGTCTCCAGT 2032
Db 367 ATATTATTTTGATAAGACTGTTTGAAATATACACAGAGGAAATTCGTGGTCTCCAGT 426

Qy 2033 TTTGCAGAACACATGATTT---GAGTCATTTTAAACATGCTAGTCTTACTTTAAGCTT 2089
Db 427 TTTGCAGAACACATGATTTTGAAGTCATTTTAAACATGCTAGTCTTACTTTAAGCTT 486

Qy 2090 GTACAAACTGCTGTATATGATGTAAACATATCTCTAGTCTGATAGTAGTTGTA 2149
Db 487 GTACAAACTGCTGTATATGATGTAAACATATCTCTAGTCTGATAGTAGTTGTA 546

Qy 2150 TTACAGCTGAACACTGCTCTCAGTGAAGGTGAGAGAGTAAGACTCTGAGTCAGAAAT 2209
Db 547 TTACAGCTGAACACTGCTCTCAGTGAAGGTGAGAGAGTAAGACTCTGAGTCAGAAAT 606

Qy 2210 CTGGGCTAAGCTGCTCACTACAGAAAGTACAAATATAAATGCAACATGATTTCT 2269
Db 607 CTGGGCTAAGCTGCTCACTACAGAAAGTACAAATATAAATGCAACATGATTTCT 666

Qy 2270 ATTTGTTTTTCTGCTGATGTTAATGATTTATTTATTTTATTTTATTTTATTTTATTTT 2329
Db 667 ATTTGTTTTTCTGCTGATGTTAATGATTTATTTATTTTATTTTATTTTATTTTATTTT 726

Qy 2330 TGTGATGAAAGAAAGTTAGGAGCCACGATATCTGAAGCTCACTATTTCAG 2381
Db 727 TGTGATGAAAGAAAGTTAGGAGCCACGATATCTGAAGCTCACTATTTCAG 778

RESULT 2
BX932340
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 923)
Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,
Croning, M.D.R., Davies, R.M., Francis, M.D., Grafham, D.V.,
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,
Tickle, C. and Willson, S.A.
Direct Submission
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chickest@bms.umist.ac.uk
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project.
This sequence is from the
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Bosch. cDNA was prepared
from RNA extracted from small intestine, normalised, and poly
A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.
Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
coli DH10B.
Location/Qualifiers
1..923
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST35022"
/clone_lib="CSEQCHN70"
/dev_stage="stage 36"

Query Match      3.6%; Score 85; DB 5; Length 923;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1610 TCCTTACAGACAGAAAGATGGCAATTTAAGCGTACTTGGAAATATGAGAAATGAGAACT 1669
Db 1 TCCTTACAGACAGAAAGATGGCAATTTAAGCGTACTTGGAAATATGAGAAATGAGAACT 60

Qy 1670 ATGAAAAATTCATGGAAGCAATGGG 1694
Db 61 ATGAAAAATTCATGGAAGCAATGGG 85

RESULT 3
BX930518
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 899)
Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,
Croning, M.D.R., Davies, R.M., Francis, M.D., Grafham, D.V.,
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,
Tickle, C. and Willson, S.A.
Direct Submission
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: chickest@bms.umist.ac.uk
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
sequencing project.
This sequence is from the
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
from a library constructed by Elizabeth Bosch. cDNA was prepared
from RNA extracted from small intestine, normalised, and poly
A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.
Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
coli DH10B.
Location/Qualifiers
1..899
/organism="Gallus gallus"
```

```
RESULT 4
BX933597
LOCUS      896 bp      mRNA      linear      VRT 02-FEB-2004
DEFINITION Gallus gallus finished cDNA, clone CHEST153f19.
ACCESSION  BX933597
VERSION     BX933597.1 GI:41634125
KEYWORDS   Gallus gallus (chicken)
SOURCE     Gallus gallus
ORGANISM   Gallus gallus
REFERENCE  1
AUTHORS    Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
            Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
            Tickle,C. and Wilson,S.A.
TITLE      Direct Submission
JOURNAL    Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
COMMENT    CB10 ISA, UK. E-mail enquiries: chickest@ms.umbist.ac.uk
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
            sequencing project.
            This sequence is from the
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
            from a library constructed by Elizabeth Bosch. cDNA was prepared
            from RNA extracted from small intestine, and poly A-trimmed.
            EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
            pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
            coli DH10B.
FEATURES   Location/Qualifiers
            source          1..896
                        /organism="Gallus gallus"
                        /mol_type="mRNA"
                        /strain="Compton Line 151"
                        /db_xref="taxon:9031"
                        /clone="CHEST153f19"
                        /dev_stage="adult"

ORIGIN
Query Match      3.4%; Score 81; DB 5; Length 899;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1614 TACAGACAGAAAGATGGCATTTAACGGTACTTGGAAATAGAGAAAAATGAGAACTATGA 1673
Db      |||||||
Qy 1674 AAAATTTCATGGAAGCAATGGG 1694
Db      |||||||
Qy 71 AAAATTTCATGGAAGCAATGGG 91
Db      |||||||

RESULT 4
BX933597
LOCUS      896 bp      mRNA      linear      VRT 02-FEB-2004
DEFINITION Gallus gallus finished cDNA, clone CHEST153f19.
ACCESSION  BX933597
VERSION     BX933597.1 GI:41634125
KEYWORDS   Gallus gallus (chicken)
SOURCE     Gallus gallus
ORGANISM   Gallus gallus
REFERENCE  1
AUTHORS    Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
            Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
            Tickle,C. and Wilson,S.A.
TITLE      Direct Submission
JOURNAL    Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
COMMENT    CB10 ISA, UK. E-mail enquiries: chickest@ms.umbist.ac.uk
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
            sequencing project.
            This sequence is from the
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
            from a library constructed by Elizabeth Bosch. cDNA was prepared
            from RNA extracted from small intestine, and poly A-trimmed.
            EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
            pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
            coli DH10B.
FEATURES   Location/Qualifiers
            source          1..896
                        /organism="Gallus gallus"
                        /mol_type="mRNA"
                        /strain="Compton Line 151"
                        /db_xref="taxon:9031"
                        /clone="CHEST153f19"
                        /dev_stage="adult"

ORIGIN
Query Match      3.3%; Score 78.8; DB 5; Length 896;
Best Local Similarity 97.6%; Pred. No. 2.7e-06;
Matches 80; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1613 CTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAATAGAGAAAAATGAGAACTATG 1672
Db      |||||||
Qy 7 CGAGACAGAAAGATGGCATTTAACGGTACTTGGAAATAGAGAAAAATGAGAACTATG 66
Db      |||||||

Qy 1673 AAAAATTCATGGAAGCAATGGG 1694
Db      |||||||
Qy 67 AAAAATTCATGGAAGCAATGGG 88
Db      |||||||
```

```
RESULT 5
AX083744/c
LOCUS      1141 bp      DNA      linear      PAT 28-FEB-2001
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION  AX083744
VERSION     AX083744.1 GI:13185472
KEYWORDS   synthetic construct
SOURCE     synthetic construct
ORGANISM   artificial sequences.
REFERENCE  1
AUTHORS    Kunst,L. and Clemens,S.
TITLE      Regulation of embryonic transcription in plants
JOURNAL    Patent: WO 011061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES   Location/Qualifiers
            source          1..1141
                        /organism="synthetic construct"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:32630"
            promoter        1..1141
                        /notes="consensus sequence of A.t., L.a., and B.n. FAEL
                        promoters"
ORIGIN
Query Match      2.7%; Score 64.2; DB 6; Length 1141;
Best Local Similarity 12.7%; Pred. No. 0.0033;
Matches 136; Conservative 379; Mismatches 550; Indels 10; Gaps 3;

Qy 830 TTTTCTGCATTATCCTTCAACATTTAAACCTCGGATCTATGGAATCAACACGTTGGG 889
Db      |||||||
Qy 1129 TTTKTKYKANNNNNNNGKDNWMDATKWSATGTAWTINAKRGATWCWYWTGTN 1070
Db      |||||||
Qy 890 TAAAAATTCATCTAGCAGACATCAACTACTGTAGGAATGGACAGAAACAGAGCATTCAC 949
Db      |||||||
Qy 1069 RRCMRTYAMRTWYTRSNANWSCATKBMWMTWKWYATKYRTAWYAMWCAMRNNNNMWCATN 1010
Db      |||||||
Qy 950 GAATGGCTATATATAGAGAATACGTAGAGGTGCTGCTGAATTTAGACTACCTATTAAA 1009
Db      |||||||
Qy 1009 GYAKSCATNNAMWYATTTAAAYAAAKWARWAGNNMRMYGAAAGKMGCMAMATMGSWA 950
Db      |||||||
Qy 1010 GAGTGAGGACACAAATGGAGAAATATCATCGCAATTTCTGTAGCTCAGCAGCTAGCTGAA 1069
Db      |||||||
Qy 949 DTAGKCNNNNNNWTDDVRWAMKAKNNNNNNAYWTACYNRAATNNNNKAWTHWKKWTHGAH 890
Db      |||||||
Qy 1070 GGTTCCTGAAACTGAACCGAG-----TTTCCCAAACTACCTGTGGATGTTTCAGTGGAT 1122
Db      |||||||
Qy 889 SKRTRHHTRTCRRTKYNNNNNNARTVYVYHHAARRWMAWWTRTNNNNNNNNNNACRNT 830
Db      |||||||
Qy 1123 CCTTCATCTCATGCTTATATGTGGAGTAGAATAGATTCTCACCAATTAAGATGGACAA 1182
Db      |||||||
Qy 829 RTWWABWKHSWCNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMH 770
Db      |||||||
Qy 1183 AGCAGAGATTGTGTTTATCTCTGGTAAATACGTTTCTCCAGTTGTATAGAGACC 1242
Db      |||||||
Qy 769 AAVTTHTDWCYKTNWNTWDMWMTWBTTTTTRMTTISTNNNNNNNNNNNNNNNNNNNN 710
Db      |||||||
Qy 1243 TCCACCAAGTATAAAGTCTCTATGCAACAAAGAAAATGTCAATACATCTCTTAGTCTCAT 1302
Db      |||||||
Qy 709 WKAYAHATNNWGCWNNNTDARETNNNTVMRREWMNTKTRWYSTTRRHVYTGATNNNN 650
Db      |||||||
Qy 1303 TATTATTTTCATTAGATAGCCGGTTTCTTACTACAACATCAAAATAGATGAACAGAAATGAA 1362
Db      |||||||
Qy 649 NNNNNNNNNNSCCTCTRMWMTMEWTMGDGMTVRKKYKWRDITTCYVDVWADSWVWYAN 590
Db      |||||||
Qy 1363 TGGGTTAGTGACTGTTTATAAAGAGAGTAATAAAG--ATACTATCATCATCTTTGAGGCAA 1420
Db      |||||||
Qy 589 WMRCDVTYTRNNYCKSYAHSYVWSSNNNMMWYRRYSARNWSSNARWTTTRNNNNMMSGBVR 530
Db      |||||||
Qy 1421 TAAAGGAGGAGAGATTTCAGCAACACAGTGTGCTTACAAGTGGAAAAACAAGTTAAACTAAA 1480
Db      |||||||
```

```
Db 529 MRWAGTMMWRHNNNNNTDTRYWWKRWABTTTVDYDMSMCAKSMRGNRNNRAWKQWAA 470
QY 1481 GTGACCCCTCCTTGACAAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCA 1540
Db 469 NNDAGADHWTYWMGNNTMMRRRAWKMMNMAWCRRAYCCNNNNNNRAACVWHKHQWRWTWK 410
QY 1541 TGTAAATGTCTTCTGATAAGCCTGTTTCATAAATCTCTTTGCAAGCTCTGCTACTTA 1600
Db 409 YMKCAACNNNNBKAMYNRVAMWYSRDTTNDMMWTSDBWHWYTVDYTMRAWNNNN 350
QY 1601 CCAGAAGCTGCGCTACAGACAGAAGATGCGATTTAAACGCTACTTGGAAATAGAGAAA 1660
Db 349 NNWRBCKTTTSMWMMMDHNTHTCTYGNNTGSAVEMASWMAAGASBNVTYNNCWRMTYM 290
QY 1661 ATGAGAACTATGAAAT-TCATGGAAGCAATGGGTAAGCTTACTTTTGTGAATGCCCT 1719
Db 289 GKTMTNNNNKAWYRTKTVAMCNRYYDTAVWTBKRYKYCYAYBWTBYBYMKGKHW 230
QY 1720 CTAAGACGAGTACCACTACGCGGAATACAACTTAAAGCTGTTCATGAACCTACCATCT 1779
Db 229 BWRERABHRSWNNWVKCRKYWVSHYHARVYKABAVGCNNWKKDRMAHHWCATNN 170
QY 1780 GGCTAACTCCTCTTGTCTGCTATTGTCCTTGGCCTTGCAATGCGCTGCACTATTTT 1839
Db 169 NMWMMYAYMHMHKGAATNNKTABRDDHBAHVKYWYRYDYWCAMCMWNAKAKVR 110
QY 1840 GAAAGAACTCTATAGAGGGAATACAGGAAGAAACATCTCTGATTTATTTCG 1894
Db 109 TAMKHMMYYTDRYVSANNTGVRMMWRMCMWYMMNNRWYRGRKYTWAMWYSM 55

RESULT 6
RATFABP
LOCUS
DEFINITION
Rat intestinal fatty acid binding protein gene, exons 1 and 2, and
a B2 repetitive element.
ACCESSION
M18080 J03465
VERSION
M18080.1 GI:204071
KEYWORDS
B2 repetitive sequence; fatty acid binding protein.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
REFERENCE
1 (bases 1 to 3013)
AUTHORS
Sweetser,D
JOURNAL
Unpublished (1987)
REFERENCE
2 (bases 917 to 1217)
AUTHORS
Sweetser,D.A., Birkenmeier,E.H., Klisak,I.J., Zollman,S.,
Sparkes,R.S., Mohandas,T., Lusis,A.J. and Gordon,J.I.
TITLE
The human and rodent intestinal fatty acid binding protein genes. A
comparative analysis of their structure, expression, and linkage
relationships
J. Biol. Chem. 262 (33), 16060-16071 (1987)
MEDLINE
88058967
PUBMED
2824476
COMMENT
Original source text: Rat (Sprague Dawley) intestinal epithelial
lining DNA, clone lambda-RIFABP.
Draft entry and computer-readable sequence for [2] kindly provided
by D.Sweetser, 19-JAN-1988.
FEATURES
source
1. .3013
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
195. .406
repeat_region
/notes="B2 repeat"
prim_transcript
1155. .>3013
/notes="FABPI mRNA (alt.) and introns"
prim_transcript
1179. .>3013
/notes="FABPI mRNA (alt.) and introns"
join(1215. .1281,2541. .2713)
CDS
/notes="fatty acid binding protein"
```

Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Mowles, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plummer, F., Poidexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savory, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 230897)  
Worley, K.C.

Direct Submission  
Submitted (11-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 230897)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23101530.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GBMN  
Center clone name: CH230-4G2  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 210222 bases at least Q40  
Consensus quality: 213326 bases at least Q30  
Consensus quality: 215519 bases at least Q20  
Estimated insert size: 218065; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_table.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_table.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 157719: contig of 157719 bp in length  
\* 157720 157819: gap of unknown length  
\* 157820 230897: contig of 73078 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..230897  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-4G2"  
1..2206  
/notes="wgs\_end extension  
clone end:Sp6"  
complement(72249..73061)  
/note="clone boundary  
clone end:Sp6  
site:EcoRI  
end sequence:BH306439"  
157820..159219  
/note="wgs contig"  
182442..183960  
/note="wgs\_contig"

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

ORIGIN  
Query Match 2.6% Score 62; DB 2; Length 230897;  
Best Local Similarity 59.3%; Pred.No. 0.0044;  
Matches 127; Conservative 0; Mismatches 80; Indels 7; Gaps 1;  
Qy 1515 TTGAGCTTTAGCCAGCCACATCATGTAATTTGCTTCTGATAGCCTGTTCTATAA 1574  
Db 156173 TTGAACCTTGAACCTTCCACATCATGTAATTTGCTTCCAGAGATAAGAAATAA 156232  
Qy 1575 TTCTCTTTGCAAGCTCTGCTACTTACCAGATC-----TGCCTACAGACAGAAGA 1627  
Db 156233 TTCTCTCTAGTGAGCAGGACCGAATCTCTCTTCTAGAGGCACACACAGCTGACATCA 156292  
Qy 1628 TGGCATTTAACGGTACTTGGAAAATGAGAAAATGAGAACTATGAAAAATTCATGAAG 1687  
Db 156293 TGGCATTTTGATGCACTTGGAAAGTAGACCGAATGAGAACTATGAAAAATTCATGAGA 156352  
Qy 1688 CAATGGGTAGCCCTTACTTTTTTGAATGCTTCT 1721  
Db 156353 AAATGGTAAAGGCTGGCTTCTCTGCTATTGCT 156386

RESULT 8  
AX083744  
LOCUS AX083744 1141 bp DNA linear PAT 28-FEB-2001  
DEFINITION Sequence 22 from Patent WO0111061.  
ACCESSION AX083744  
VERSION AX083744.1 GI:13185472  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1  
AUTHORS Kunst, L. and Clemens, S.  
TITLE Regulation of embryonic transcription in plants  
JOURNAL Patent: WO 011061-A 22 15-FEB-2001;  
UNIVERSITY OF BRITISH COLUMBIA (CA)  
FEATURES  
source  
Location/Qualifiers  
1..1141  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
1..1141  
promoter

ORIGIN	/note="consensus sequence of A.t., L.a., and B.n. FAbI promoters"									
Query Match	2.4%	Score	57.8	DB	6	Length	1141			
Best Local Similarity	10.9%	Pred. No.	0.075							
Matches	82	Conservative	263	Mismatches	407	Indels	0	Gaps	0	
QY	1565	TGTCATATAATCTCTTTGGCAAGCTCTGCTACTTACAGAAAGTCTGCCTACAGACAGAA	1624							
Db	201	DRDWSBKRMNYGMBWKNWSYDVVYVWVDDMKRVRRVWTRGRMRVYVWAWBTAHR	260							
QY	1625	AGATGGCATTAAACGGTACTTTGGAAATAGAGAAATAGCAACTAGAAAATTCATGG	1684							
Db	261	RYVNGWTBAYARRWTNNNNNAKACRKAKYWGWNBRABVNSTCTTWSKTKTKVRTSC	320							
QY	1685	AAGCAATGGGTAAGCCTTCTTTTGAATGCCCTTCTAAAGCAGGATACCACTACGGCG	1744							
Db	321	WANNCRAGDANKDKHKWKWSAAMGVYNNNNNNNTYKKARHBARWDVWHSWKKWHAN	380							
QY	1745	GAATACAAACTTAAGCTTTCATGAACACTACATCTGGCTAACCTGTCTCTTTGTGTCTGC	1804							
Db	381	AHYSRKKTBYKRTKVMVNNNGTTTMMKRWAWYMKMDMBGTYNNNNNGRYYGWTK	440							
QY	1805	TATTTTGCCTTGCACATTGCCCTGCCTTATTTTGAAGAGACTCTATAGAGGGGAATAC	1864							
Db	441	NKMWYTYKWKANNCKRWDHKTCTNNNTTWKMKTYWNNCTWKSNTNGSHRBAAYV	500							
QY	1865	AAGGAAGAAACATCTCGATTTTATTTGCAATTCGCGATAATCTTATGCAATTTAGCTAAT	1924							
Db	501	TWYMWRRYAHANNNDWYWKACTWYKVBVCSKWNYYAAMWTKSSWNTSRYYRWKT	560							
QY	1925	CCAGTAGGCATTCAGCAGAAATTAATAGAAATATATGTAAGGAATATTTTGA	1984							
Db	561	NNSWRWSDTRSMDRANRYARABHYGKNTRWBSHSTWBHBRAGAAHYWMBMBAK	620							
QY	1985	TAAGACTGTTTGAATAATTACAGAGGGGAAATTTGCTGCTCCAGATTTTGCGAGAACAC	2044							
Db	621	HCWKAWYKAKYAGAGSGNN	2104							
QY	2045	ACATGATTGAGTCATTTTAACATGCTAGTGGCTTACTTTAAGCTTGTACAAACTGCCTGT	2104							
Db	681	BAANNAYYTHANNWGWGNNATDTRTMMKNNNNNNAGTWKNNNNNNNAKNAKNAKNA	740							
QY	2105	AATATGATGATAAACAATACTATCCATGTTGGATAGTAGTCTTATATACAGGCTGAACAC	2164							
Db	741	AVDAKCHRWANKWNRGWHDAAABTTDKRNNGAYTKYTTNNNNNTYRGVVNTAARD	800							
QY	2165	TGCTCAGTGAAGGTGGAGAGTAGTAAGACTCTGAGTCAGAAATCTGGGCTAAGCTCCC	2224							
Db	801	GWANN	2284							
QY	2225	TCAACTACAGAAAAGTCACAATAAATAAGCAACATGATGTTCTATTTTGTGTTTCTCT	2284							
Db	861	WRBAYTNNNNNNRMAYYAYADDDYAYYMSDTCDAWMKWDAKNNNNATTYNRGTAWRTNN	920							
QY	2285	GCTTGATGTTAATGATTATATATATTTT	2316							
Db	921	NNNNWTKYKBAAAMNN	952							
RESULT 9										
MUSFABPI										
LOCUS	MUSEFABPI	5039 bp	DNA	linear	ROD 27-APR-1993					
DEFINITION	Mouse Fabpi gene, exons 1-4.									
ACCESSION	M65033									
VERSION	M65033.1	GI:193218								
KEYWORDS										
SOURCE	Mus musculus (house mouse)									
ORGANISM	Mus musculus									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
REFERENCE	1 (sites)									

AUTHORS	Green,R.P., Cohn,S.M., Sacchettini,J.C., Jackson,K.E. and Gordon,J.I.									
TITLE	The mouse intestinal fatty acid binding protein gene: nucleotide sequence, pattern of developmental and regional expression, and proposed structure of its protein product									
JOURNAL	DNA Cell Biol. 11 (1), 31-41 (1992)									
MEDLINE	92153306									
PUBMED	1739433									
REFERENCE	2 (bases 1 to 5039)									
AUTHORS	Green,R.P., Sacchettini,J.C., Jackson,K.E., Cohn,S.M. and Gordon,J.I.									
TITLE	The mouse fatty acid binding protein gene (Fabpi): nucleotide sequence analysis and predictions of the structure of its protein product									
JOURNAL	Unpublished (1991)									
COMMENT	Original source text: Mus musculus (strain DBA/2J) Adult Liver DNA.									
FEATURES	Location/Qualifiers									
source	1..5039									
	/organism="Mus musculus"									
	/mol_type="genomic DNA"									
	/strain="DBA/2J"									
	/db_xref="taxon:10090"									
	/tissue_type="Liver"									
	/dev_stage="Adult"									
misc_feature	467..547									
	/note="Domain I: homology with human I-FABP promoter region'; putative"									
repeat_region	524..537									
	/note="one of two repeats homologous to consensus repeat found in number of other lipid binding proteins which are expressed in small intestine'; putative"									
misc_feature	727..817									
	/note="Domain II: homology to human I-FABP promoter region'; putative"									
misc_feature	870..966									
	/note="Domain III: homology with human I-FABP promoter region'; putative"									
repeat_region	884..897									
	/note="Second of two repeats homologous to consensus repeat'; putative"									
TATA_signal	937..942									
exon	967..1069									
	/note="putative"									
	/label=Exon1									
exon	2326..2497									
	/note="putative"									
	/label=Exon2									
exon	4291..4398									
	/note="putative"									
	/label=Exon3									
exon	4818..5039									
	/note="putative"									
	/label=Exon4									
ORIGIN										
Query Match	2.4%; Score 56.6; DB 10; Length 5039;									
Best Local Similarity	58.5%; Pred. No. 0.11;									
Matches	121	Conservative	0	Mismatches	79	Indels	7	Gaps	1	
QY	1516	TGAGCTTTAGCCAGCCACATCATCTGTAAATTCCTTCCTGATAAGCTGTTTCATTAAT	1575							
Db	884	TGAATCTTGAATCTCCATCATCACAGTATGAATTTGGTTCGAAAGATAAGAAATAGATAAT	943							
QY	1576	TCTCTTTGCAAGCTCTGCTACTTTACCAGAAAGTC-----TGCTACAGACAGAAAGAT	1628							
Db	944	TCTCTCTAGTGGACAGGACTGGACCTCTGCTTCTTAGAGACACACACAGCTGAGATCAT	1003							
QY	1629	GGCATTTAACCGTACTTGGAAAAATAGAGAAAAATGAGAACTATGAAAAATTCATGGAAGC	1689							
Db	1004	GGCATTTCACGGCAGCTGGAAAGTAGACCGGACGAGAACTATGAAAAAGTTTCATGGAGAA	1063							
QY	1689	AATGGTAAAGCTTACTTTTTTGAATG	1715							



Db

1064

AATGGGTAAGCACTGGCTTCCTGCTG

1090

|||||

|||||

|||||

|||||

|||||

PFMAL4P2

234081 bp

DNA

linear

INV 29-JAN-2003

LOCUS

Plasmodium falciparum MAL4P2.

DEFINITION

AL035475 AL844503

ACCESSION

AL035475.7

VERSION

GI:23498187

KEYWORDS

SOURCE

Plasmodium falciparum 3D7

ORGANISM

Plasmodium falciparum 3D7

REFERENCE

1

AUTHORS

Hall N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C.O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., Mclean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A., Rabinowitz, E., Rajandream, M.A., Rutter, S., Rutherford, K.M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and Barrell, B.G.

TITLE

Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13

JOURNAL

Nature 419 (1996), 527-531 (2002)

MEDLINE

2255708

PUBMED

12368867

REFERENCE

2

AUTHORS

Devlin, K., Pain, A., Berriman, B., Hall, N., Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.

TITLE

Direct Submission

JOURNAL

Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT

On Oct 3, 2002 this sequence version replaced gi:6562702.  
For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).

FEATURES

Location/Qualifiers

source

1. .234081  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="4"  
complement (1. .3390)  
/gene="PF00320c"  
complement (<1. .3390)  
/gene="PF00320c"  
/note="Possible distant homologue of reticulocyte binding proteins of P. falciparum  
Similar to Plasmodium falciparum hypothetical 136.1 kDa protein mal4p2.01 SWALL:Q9U0M1 (EMBL:AL035475) (1130 aa) fasta scores: E(): 0, 99.46% id in 1130 aa, and to Plasmodium falciparum reticulocyte binding protein 2 homolog a SWALL:Q9BR46 (EMBL:AF312916) (3130 aa) fasta scores: E(): 3.46-23, 22.92% id in 2753 aa, and to Plasmodium falciparum reticulocyte binding protein 2 homolog b rbp SWALL:AAQ02259 (EMBL:AF312917) (3254 aa) fasta scores: E(): 5.8e-21, 21.63% id in 2893 aa"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB62842.1"  
/db\_xref="GI:6562703"  
/db\_xref="TrEMBL:Q9U0M1"

gene

CDS

gene

translation="MDQKNLILICGKNDISKEETKEHANVLLGCAKTEKEENLVDD KIYNKSNLINIINNKKDLMKTTDDVKNDKTLKDDTLERINQKVVKTKKNHLINEREN IHEDINQNGKSPDDNKSPDDNKSPDDNKSPDDNKSPDDNKSPDDNKSPDDNKSPDDNK SKPDNDKSNPDNKCIMNDKNTGDIKKNKIDNVFINRCSNYSYRVVINTNRIYNYV ENNIRQKGFHFVKYKNYIHYAEKVMNIKEIKRKYBENYHFAKSYLCHDLLANIKKOT IMFVKGKFDENRYNNTLCEHKNKDNKIFNEINGLRSILSEDKTNMBIKODEYYKITL QTHQCFKDKENVPVDIMNIHKDVTYDNATNINKBDNNYLDMDHIMKEMSNKTEG SYLYYMNNDMPESKNVYKIYIMKSEGNVTRHMKDIDEGRCIYNKCSKNYSINHY SGINICLGNVKGEDHTFVMPYDKKWEIRKSTNNNNINRDKRNTHRSLKNVR IINIKHKNISYICKNLKGGKYYERRMLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNL KYILOSVPNKLITNKNIMDYNKINTIIGKEKEKEKDNKINIKHTKHSKHNKNNNNN NKEKPDNDINIVKMKKRTYIFPINKTYKLFIHNYKNISKFENLARKHFVADYDIF KLSYVITLKEKIDLCFLYMDWALEILDKNLDIHKENKLIQNHLCTINLREIENI KROLFLEYSKHSRFLCKSLHEWETNKYIFLNKKKENTONDIHSLVLINISIKTE EQMDIHKYMYTHKYTHYKYKSPAYGKTAQMNENLEIMKCRISISQIENLEKKIQ REIMFTEOKRMEKKEKIFIEBKKIQKQYIDENKKNTERKNTLIDUEQINHEQT KNNTNKELEKEIKNFENEKITIEKQNSNDKITLEKEIQNTIRNEKMTLEKEIQ NISNDKITIEKIKNFRNDKITLEKIKNFRNDKMTLEKIKNFSNDKITLEKIQNI RNEKITIEKIQNISDKVTIEKIQIRNDEMTLEKIQNF"

gene

5644. .6063  
/gene="PF00325w"

CDS

5644. .6063  
/genes="PF00325w"  
/notes="Similar to Plasmodium falciparum very hypothetical protein, Mal4p2.02 mal4p2.02 SWALL:Q9U0M0 (EMBL:AL035475) (139 aa) fasta scores: E(): 1.4e-47, 98.57% id in 140 aa"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB62843.1"  
/db\_xref="GI:6562704"  
/db\_xref="TrEMBL:Q9U0M0"  
/translation="MYRNIIKYFLPTWKRTRYHRSKQNNLSNKNIKDPYSLYKSS MYGNFKILPNKATKREYEILKTSNNTSYTSYPPNINITLTPPESSKMYENR KYIMKYKNVEYIPIKRLTYKNAKSNNTWNTYIRMEK"  
join (6794. .6857,6942. .7075,7272. .7383,7566. .7663, 7792. .8064,8151. .8357,8454. .8588,8638. .9609,9740. .10018, 10170. .10250,10340. .10378,10551. .10766)  
/gene="PF00330w"  
join (6794. .6857,6942. .7075,7272. .7383,7566. .7663, 7792. .8064,8151. .8357,8454. .8588,8638. .9609,9740. .10018, 10170. .10250,10340. .10378,10551. .10766)  
/gene="PF00330w"  
/notes="There is slight modification in the gene model in the last exon from submitted (SWALL:Q9U0L9 (EMBL:AL035475)) Plasmodium falciparum hypothetical 103.2 kDa protein sequence  
Similar to Plasmodium falciparum hypothetical 103.2 kDa protein mal4p2.03 SWALL:Q9U0L9 (EMBL:AL035475) (853 aa) fasta scores: E(): 0, 100% id in 850 aa"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD49159.1"  
/db\_xref="GI:23498188"  
/db\_xref="TrEMBL:Q81W4"  
/translation="MNVEDRYKLKEKYKEIKQNDILKKAILEYKDKLEKKNDI LNDKNQLQKNTLTNSLEQKKKSGWNTNMLTKNSRENTHKSVAFELMKIJE NLTNKKIEDLHOEKSYLEKLLKDKHIEKREKELTGTUNDLVNSHEKQIKAD EQYEEIKKLTIEENQKYSQOIKKKEESLDYIRNISTMLHKNKPKYMINDFQIPFYQA YDTKGSYIQKEILQNEERLANLYLSIQVYVLYKVKELFLQKNLDSMYNKKKE KLEQOKYSKMDLKNICQKVLIEEIDTAEKFKIYAEKEDKELMEVYFLNIYNI QMYTHVNIPIIKKIRLKVNVYLCIEEYVFPFSINNKKVFLRNVIQQPRLKXE LVTKIITLTIISICPNNERIILENPKNTIKKENNKIINHEKRNDEKYNIS SFDWITHTIYDDDKLQFVNKQYENKSLYNPFIVEKKIPFIFLQKEKIMLSQFND FTYQTKNNHILFLIKKIQILOYIPQSLFKLSYSIVRLCEINRFGNYINNSPQI TNLDDKTNBFIKS IENFYTHIHNIIYTKLSYINHYTPYLYKQHYIINNITFNKA KNYVYDQLQKNLQDNFKQNLQNLKSKKNKKLKTLTNFNINLLEQELSL KNOYLNLSDNKEDQYSNDNNNIOTYAEKIYNFISHTDEGKKFNKEKQILEAVI SSCIKINLNLEIQNNKHILBELQNLFNKENEIKNLNTQIDTYKEEVEVNIHKYEQ MCAHFYFPHFFHF"  
complement (10868. .19509)  
/gene="PF00335c"

gene



```
prim_transcript 1028. .>4393
/map="4q28-q31"
/note="FABPI mRNA (alt.) and introns"
```

```
prim_transcript 1053. .>4393
/notes="FABPI mRNA (alt.) and introns"
CDS
join(1089..1155,2350..2522,3546..3653,4098..4148)
/notes="fatty acid binding protein"
/codon_start=1
/protein_id="AA52417.1"
/db_xref="GI:182352"
/translations="MAPDSTWVKVDRSENVDKFKMGVNIIVKRLAAHDNKLITIOE
GNKFTYKESAFPNIEVPELGVTFFYNLADGTGLGTSLEGNKLIKFKEKTDNGNE
LNIVRIIGDELQVTYYGVEAKRIFKKD"
gene 1089..1155
exon <1089..1155
/genes="FABP2"
/notes="fatty acid binding protein; G00-119-127"
intron 1156..2349
/notes="FABPI intron A"
exon 2350..2522
/number=2
intron 2523..3545
/notes="FABPI intron B"
exon 3546..3653
/number=3
intron 3654..4097
/notes="FABPI intron C"
exon 4098..4148
/notes="fatty acid binding protein"
repeat_region 4466..4742
/notes="Alu repeat"
repeat_region 4466..4472
/notes="5' direct repeat"
repeat_region 4736..4742
/notes="3' direct repeat"
ORIGIN 156 bp upstream of HindIII site; chromosome 4q28-q31.
Query Match 2.1%; Score 49.8; DB 9; Length 5204;
Best Local Similarity 59.2%; Pred. No. 3;
Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
QY 1516 TGAGCTTTAGCCAGCACATCATGTAAATGCTTTCTGTGATAAGCCTGTTCAATAAT 1575
Db 971 TGAACCTTAAGCTTCCACATGATGAGTTGGTTCAAGATAAGAAATAATAATAAT 1030
QY 1576 TCTC-----TTTGCAGAGCTCTGCTACTTACAGAGTCTGCTTACAGACAGAAAGA 1627
Db 1031 TCTCGCCCAAGGACAGACCTGAAATCTTAGCTGCCTAGAGGCTGACT-CAACTGAAATCA 1089
QY 1628 TGGCATTTAACGGTACTTGGAAATAGAGAAATAGAGAACTATGAAAAATTCATGGAAG 1687
Db 1090 TGGCGTTTGACACACTTGGAGGTAGACCGGAGTGAAACTATGACAACTTCAATGGA 1149
QY 1688 CAATGGGTAGCCTTACTTTTGAATGCTT 1718
Db 1150 AAATGGGTAAGACTTTATTCTTTGTGGCT 1180
RESULT 15
AC008077/c 20000 bp DNA linear HTG 20-JUL-1999
LOCUS Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 20
DEFINITION unordered pieces.
ACCESSION AC008077
VERSION AC008077.1 GI:5540103
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 20000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
```

```
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1146: contig of 1146 bp in length
1147 10118: gap of unknown length
1148 10119: contig of 1190 bp in length
1149 20280: gap of unknown length
1150 21488: contig of 1208 bp in length
1151 30460: gap of unknown length
1152 31634: contig of 1174 bp in length
1153 40606: gap of unknown length
1154 41979: contig of 1373 bp in length
1155 50951: gap of unknown length
1156 52049: contig of 1098 bp in length
1157 61021: gap of unknown length
1158 62105: contig of 1084 bp in length
1159 71077: gap of unknown length
1160 72391: contig of 1314 bp in length
1161 81364: gap of unknown length
1162 82566: contig of 1203 bp in length
1163 91538: gap of unknown length
1164 92500: contig of 1362 bp in length
1165 101871: gap of unknown length
1166 103074: contig of 1203 bp in length
1167 112045: gap of unknown length
1168 113302: contig of 1257 bp in length
1169 122273: gap of unknown length
1170 123778: contig of 1505 bp in length
1171 132749: gap of unknown length
1172 134818: contig of 2069 bp in length
1173 143789: gap of unknown length
1174 145506: contig of 1717 bp in length
1175 154477: gap of unknown length
1176 156731: contig of 2254 bp in length
1177 165702: gap of unknown length
1178 167355: contig of 1653 bp in length
1179 176326: gap of unknown length
1180 178351: contig of 2025 bp in length
1181 187322: gap of unknown length
1182 188990: contig of 1668 bp in length
1183 197961: gap of unknown length
1184 200000: contig of 2039 bp in length.
FEATURES
Location/Qualifiers
source
1..200000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
ORIGIN
Query Match 2.1%; Score 49.8; DB 2; Length 200000;
Best Local Similarity 59.2%; Pred. No. 1.7;
Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
QY 1516 TGAGCTTTAGCCAGCACATCATGTAAATGCTTTCTGTGATAAGCCTGTTCAATAAT 1575
Db 10910 TGAACCTTAAGCTTCCACATGATGAGTTGGTTCAAGATAAGAAATAATAATAAT 10851
QY 1576 TCTC-----TTTGCAGAGCTCTGCTACTTACAGAGTCTGCTTACAGACAGAAAGA 1627
Db 10850 TCTCCGCCCAAGGACAGACCTGAAATCTTAGCTGCCTAGAGGCTGACT-CAACTGAAATCA 10792
```

Qy	1628	TGGCATTTAACGGTACTTGGAAAAATAGAGAAAAATGAGAACTATGAAAAATTCATGGAAG	1687
Db	10791	TGGCGTTTTCAGACACTTGGAGGTAGACCGGAGTGAACACTATGACAAGTTCATGGAAG	10732
Qy	1688	CAATGGGTAAAGCCTTACTTTTTCGAATGCCT	1718
Db	10731	AAATGGGTAAAGACTTTATTCTTTTGGCT	10701

Search completed: January 14, 2005, 15:42:20  
Job time : 10555.8 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 05:22:03 ; Search time 1240.89 Seconds  
(without alignments)  
10072.512 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agctctctgcgcagaaagg.....atctgaagctcactattcag 2381

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4134886 seqs, 2624710521 residues  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_23Sep04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2381	100.0	2381	11	ADL90127
2	336	14.1	336	11	ADL90128
3	50.6	2.1	4590	5	AAH24065
4	49.8	2.1	5204	6	ABK83946
5	49.8	2.1	5204	12	ADJ74588
6	49.8	2.1	5204	12	ADJ78215
7	49.2	2.1	2000	8	ADA71938
8	47.2	2.0	337	8	ABX40871
9	47	2.0	2636	4	ABL15880
10	46.8	2.0	662	10	ADD29587
11	46.6	2.0	610	4	ABL15953
12	46.6	2.0	9742	6	ABL70479
13	46.2	1.9	6222	6	ABL32692
14	45	1.9	10652	6	ABN80056
15	44.4	1.9	2000	8	ADA71938
16	44.2	1.9	515	3	AAA16469
17	43.6	1.8	14023	6	ABL34104
18	43.2	1.8	7143	6	ABL32983
19	42.6	1.8	513	12	ACH79811
20	42.6	1.8	611	3	AAA16371
21	42.6	1.8	2844	10	ABT31903

C	22	42.6	1.8	5501	6	ABL32055	Human imm
	23	42.4	1.8	230	2	AAH11681	Human bia
	24	42.4	1.8	230	2	AAH12923	Human bia
C	25	42.4	1.8	452	3	AAA77858	CDNA enco
C	26	42.4	1.8	452	4	AAI28596	Colon tum
C	27	42.4	1.8	452	8	ABZ32782	Human col
C	28	42.4	1.8	461	4	AAI10372	Human bre
C	29	42.4	1.8	5908	4	AAH45387	Chemical
C	30	42.4	1.8	5908	6	ABK28232	DNA trans
	31	42.2	1.8	453	6	ABL67809	Oesophagu
	32	42.2	1.8	453	6	ABL62188	Colon ade
	33	42.2	1.8	6621	6	ABL32912	Human imm
	34	42	1.8	9095	6	ABQ67061	Human ang
C	35	41.8	1.8	10048	6	ABL70313	Chemical
C	36	41.8	1.8	10048	6	AAH61251	Human gen
	37	41.8	1.8	137870	10	ADG89426	Human mat
C	38	41.6	1.7	900	4	AAH34536	Human col
C	39	41.6	1.7	3639	6	ABL62356	Colon ade
C	40	41.6	1.7	3639	12	ADJ74884	Marker ge
C	41	41.6	1.7	3719	9	ADA10896	Human CDN
	42	41.4	1.7	447	8	ABX50540	Bovine ES
C	43	41.4	1.7	6222	6	ABL32693	Human imm
	44	41.4	1.7	6782	6	ABL32777	Human imm
	45	41.4	1.7	11260	4	AAH45315	Chemical

ALIGNMENTS

RESULT 1  
ADL90127  
ID ADL90127 standard; DNA; 2381 BP.  
XX  
AC ADL90127;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region.  
XX  
KW Chicken; ds; intestinal fatty acid binding protein; iFABP;  
KW gut specific promoter; transgenic.  
XX  
OS Gallus gallus.  
XX  
PN US2003177516-A1.  
XX  
PD 18-SEP-2003.  
XX  
PF 14-MAR-2002; 2002US-00099563.  
XX  
PR 14-MAR-2002; 2002US-00099563.  
XX  
PA (HORS/) HORSEMAN N D.  
PA (PRAT/) PRATT S L.  
XX  
PI Horseman ND, Pratt SL;  
XX  
PP WPI; 2003-898653/82.  
XX  
DR New nucleic acid molecule comprising an isolated avian gut-specific gene  
PT expression control region, useful for regulating heterologous nucleic  
PT acids in transgenic avians, and for generating transgenic birds.  
XX  
PS Claim 1; SEQ ID NO 1; 28pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising an isolated  
CC avian gut-specific gene expression control region appearing as  
CC ADL90127 (Chicken intestinal fatty acid binding protein, iFABP, gene, 5'  
CC region or ADL90128 (Chicken iFABP promoter) or its degenerate variant  
CC Also included are a recombinant DNA molecule comprising an isolate avian  
CC gut-specific gene expression control region operably linked to a nucleic  
CC acid insert encoding a polypeptide, an expression vector that integrates  
CC into a host cell (and comprising the isolated avian gut-specific gene

expression control region), expressing a heterologous polypeptide in a host cell (by transfecting a eukaryotic cell with the recombinant DNA molecule, and culturing the transfected cell in a medium suitable for expression of a heterologous polypeptide under the control of an avian intestinal fatty acid binding protein (IFABP) or cp35 gene expression control region encoded by the recombinant DNA molecule), a eukaryotic cell transformed with the expression vector (or its progeny, which expresses a heterologous polypeptide) and a transgenic avian having a heterologous polynucleotide sequence comprising the nucleic acid insert. The nucleic acids are useful for regulating heterologous nucleic acids in transgenic avians, as probes in nucleic acid hybridisation assays for detecting the IFABP gene expression control region, and for generating transgenic birds. The present sequence is the Chicken intestinal fatty acid binding protein, IFABP, gene, 5' region.

Query Match	100.0%	Score 2381;	DB 11;	Length 2381;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2381;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AGCTTCTCGCCAGAAAAGGCTGTGGGGTCTCTGTCCCTCACACAGCTTAAGCAAAATC	60	
Db	1	AGCTTCTCGCCAGAAAAGGCTGTGGGGTCTCTGTCCCTCACACAGCTTAAGCAAAATC	60	
Qy	61	CCCAAGTTCAAAACGTCGGCTGTAAAGGAGATGGCTCACCTTCAAATGAAGTGAATTATG	120	
Db	61	CCCAAGTTCAAAACGTCGGCTGTAAAGGAGATGGCTCACCTTCAAATGAAGTGAATTATG	120	
Qy	121	AAATAAATCAATAAACGAGCTCTGTGGCAGATCAGAGATAACCTCTGCTGGGACAAAATT	180	
Db	121	AAATAAATCAATAAACGAGCTCTGTGGCAGATCAGAGATAACCTCTGCTGGGACAAAATT	180	
Qy	181	CTTAAAGTGATAGGTAGAACAGAGAGTGTTCAACTTAAATGACATAGATTGTACCACA	240	
Db	181	CTTAAAGTGATAGGTAGAACAGAGAGTGTTCAACTTAAATGACATAGATTGTACCACA	240	
Qy	241	TTTGATCTTCTAGGAGACAAAGGGTCTGGAAAACAAATTAATTTCTGGTGACAGTCAGT	300	
Db	241	TTTGATCTTCTAGGAGACAAAGGGTCTGGAAAACAAATTAATTTCTGGTGACAGTCAGT	300	
Qy	301	AGCAGCCTGTTTTGGGTGCAACTACAGCACTTTGTGTGCAACAATAACAACTCTAAGTTG	360	
Db	301	AGCAGCCTGTTTTGGGTGCAACTACAGCACTTTGTGTGCAACAATAACAACTCTAAGTTG	360	
Qy	361	TTTTCTTTTCTCTTCTTAACTTCTGTACAGTCTAAAGGTGAAGAGTAGCTATTTCAGT	420	
Db	361	TTTTCTTTTCTCTTCTTAACTTCTGTACAGTCTAAAGGTGAAGAGTAGCTATTTCAGT	420	
Qy	421	TACTTCCCTCTGCATCCTCTTAGCCAGATTAGCATTTGATTTCAAATTAACCTGAGTGGGA	480	
Db	421	TACTTCCCTCTGCATCCTCTTAGCCAGATTAGCATTTGATTTCAAATTAACCTGAGTGGGA	480	
Qy	481	ATGGAAAGCCACACTATTTTGGTATCACACGAAAGTTCTAAATTTATAGTTATPACTTCA	540	
Db	481	ATGGAAAGCCACACTATTTTGGTATCACACGAAAGTTCTAAATTTATAGTTATPACTTCA	540	
Qy	541	GTAAAACCTTTTGGTCGAGTCTGGAGAGAAAAAGAGATTATGATAACACGAGACTAGTA	600	
Db	541	GTAAAACCTTTTGGTCGAGTCTGGAGAGAAAAAGAGATTATGATAACACGAGACTAGTA	600	
Qy	601	AAATTCAATTAGTTAGAGCCACCTGTTATCTGTGTGATAGCAACATTCAATTTCAGCAT	660	
Db	601	AAATTCAATTAGTTAGAGCCACCTGTTATCTGTGTGATAGCAACATTCAATTTCAGCAT	660	
Qy	661	TCAGGATTTTACATTTTGTGAAGCTAAATAGACAGACAGATTTGGTGCCTGCCATAGGAAACAGA	720	
Db	661	TCAGGATTTTACATTTTGTGAAGCTAAATAGACAGACAGATTTGGTGCCTGCCATAGGAAACAGA	720	
Qy	721	CTAACTATAACTCTGAGTTTAGTACAAGCAGATTTAGCACAGCAAAATTTGCTCAGTTTC	780	
Db	721	CTAACTATAACTCTGAGTTTAGTACAAGCAGATTTAGCACAGCAAAATTTGCTCAGTTTC	780	

Qy	781	AAGTAGCACTACTTTGTGGGGAAGAGAGCTGAGCCAGTGTGTCTCATTTTCTGCATT	840
Db	781	AAGTAGCACTACTTTGTGGGGAAGAGAGCTGAGCCAGTGTGTCTCATTTTCTGCATT	840
Qy	841	ATCCTTCAACATTTTAAACCTGGGATCTATGGAAATCAAAACACGTCGGTAAATTCAC	900
Db	841	ATCCTTCAACATTTTAAACCTGGGATCTATGGAAATCAAAACACGTCGGTAAATTCAC	900
Qy	901	TAGCAGCACATCAACTACTGTAGGAATGGACAGAAAACAGAGCATTCAC	960
Db	901	TAGCAGCACATCAACTACTGTAGGAATGGACAGAAAACAGAGCATTCAC	960
Qy	961	AATATAGAGATAGCTAGAGAGTGTCTGGAATTTTAGACTACCTATTAAAGAGTGAGACA	1020
Db	961	AATATAGAGATAGCTAGAGAGTGTCTGGAATTTTAGACTACCTATTAAAGAGTGAGACA	1020
Qy	1021	CGAATGGAGAAATATCATCGCAATTTCTGTAGCTCAGCACTAGACTCGAAGGTTTCTGAAA	1080
Db	1021	CGAATGGAGAAATATCATCGCAATTTCTGTAGCTCAGCACTAGACTCGAAGGTTTCTGAAA	1080
Qy	1081	CTGAACCGAGTTTCCAAAACATACCTGTGGATGTTTCAGTGGATCCTCATCTCATGCTTAT	1140
Db	1081	CTGAACCGAGTTTCCAAAACATACCTGTGGATGTTTCAGTGGATCCTCATCTCATGCTTAT	1140
Qy	1141	TATGTGGAGTAGAATAGATTCTCACCAAAATTTAGAAATGGACAAAACAGAGATTTCTGTTTT	1200
Db	1141	TATGTGGAGTAGAATAGATTCTCACCAAAATTTAGAAATGGACAAAACAGAGATTTCTGTTTT	1200
Qy	1201	ATCTGTTGGGTAATAACGTTTTTCCAGTTGTATPAAAGACCCCTCCACAGTATAAAGTC	1260
Db	1201	ATCTGTTGGGTAATAACGTTTTTCCAGTTGTATPAAAGACCCCTCCACAGTATAAAGTC	1260
Qy	1261	CTATGCAACAAAGAAATGTCAATACATTTCTCTTAGTCTCATTTATTTTTCATTAGATA	1320
Db	1261	CTATGCAACAAAGAAATGTCAATACATTTCTCTTAGTCTCATTTATTTTTCATTAGATA	1320
Qy	1321	GCCGGTTTTTTTACTACAACCTCAAATAAGATGAACAGAAATGAATGGGTTAGTGACTGTTTA	1380
Db	1321	GCCGGTTTTTTTACTACAACCTCAAATAAGATGAACAGAAATGAATGGGTTAGTGACTGTTTA	1380
Qy	1381	TAAAGAGAGTATAAAGATACTATCATTTTGGAGCAATAGGGAGGAGAGATTCAG	1440
Db	1381	TAAAGAGAGTATAAAGATACTATCATTTTGGAGCAATAGGGAGGAGAGATTCAG	1440
Qy	1441	CAAAACAGTGTCTTACAAGTGGAAAAACAGTAAACTAAAAGTCAACCCCTCTTGACAA	1500
Db	1441	CAAAACAGTGTCTTACAAGTGGAAAAACAGTAAACTAAAAGTCAACCCCTCTTGACAA	1500
Qy	1501	GATCAATGCCACAGTTGAGCTTTAGCAGCCACATCATGTAAATTTGTTTCTCTGATA	1560
Db	1501	GATCAATGCCACAGTTGAGCTTTAGCAGCCACATCATGTAAATTTGTTTCTCTGATA	1560
Qy	1561	AGCTGTTTCATAAATCTCTTTGCAAGCTCTGCTACTTACCAGAGTCTGCCTACAGAC	1620
Db	1561	AGCTGTTTCATAAATCTCTTTGCAAGCTCTGCTACTTACCAGAGTCTGCCTACAGAC	1620
Qy	1621	AGAAAGATGGCATTTAACCGTACTTGGAATAATAGAAAAATGAGAAATTC	1680
Db	1621	AGAAAGATGGCATTTAACCGTACTTGGAATAATAGAAAAATGAGAAATTC	1680
Qy	1681	ATGGAAGCAATGGTAGCCCTTACTTTTTTGAATGCTTCTTAAACAGAGATACCACTAC	1740
Db	1681	ATGGAAGCAATGGTAGCCCTTACTTTTTTGAATGCTTCTTAAACAGAGATACCACTAC	1740
Qy	1741	GGCGGAATCAAACTTTAAGCTGTTTCATGAATACCACTGGCTTAACCTGTTTGT	1800
Db	1741	GGCGGAATCAAACTTTAAGCTGTTTCATGAATACCACTGGCTTAACCTGTTTGT	1800
Qy	1801	CTGCTATTTTGGCCCTTGCACATTTGCCCTGCACATTTTAAAGAGACTCTATAGAGGGGA	1860
Db	1801	CTGCTATTTTGGCCCTTGCACATTTGCCCTGCACATTTTAAAGAGACTCTATAGAGGGGA	1860
Qy	1861	ATACAAGGAAGAAAAACATTCATTTTATTTTTCATTTGGATGAATCTTATGCAATTTAGCT	1920











XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri P, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX MPI; 2003-175290/17.  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
  
Query March 2.1k; Score 49.2; DB 8; Length 2000;  
Best Local Similarity 10.84; Pred. No. 0.24; Mismatches 1; Gaps 1;  
Matches 83; Conservative 322; Indels 1;  
  
QY 839 TTATCTTCAACATTTAAACCTGGATCTATGGAATCAACACAGCTGGTGAATAATCA 898  
DB 291 KKSYSWYCYMYRGGRGATRYGMYGSMYMMYKMYWYRGYKMGKRWAGRMWR 350  
QY 899 CTTAGCAGCACATCACTACTGTAGGAATGGACAGAACAGACGATTCACCTGAATGGCT 958  
DB 351 SMCWSKACYMYRWRMRWMTRRRWAKSSRTSRKKKCMKMKYKRMGRYMRMSCK 410  
QY 959 ATAATATAGAGATACGTAGAGAGTCTCGAATTTAGACTACCTATTAAGAGTGAGGA 1018  
DB 411 RARMKRCRSGRAWKMGCGMTCRMKSYGMWRKWSKMRASKYKMSRYMRWRKCKSR 470  
QY 1019 CACGAATGGAGATATCATCCAAATTTCTGTAGCTCAGACTAGACTCGAAGGTTCTGA 1078  
DB 471 TTMGKTRGMMGTMGRCRYKRGSKMRKCRRRGRNRYMRWRYMYRYMYRMYCARK 530  
QY 1079 AACTGAACCGAGTTTCCCAAACTACCTGTGGATGTTTCAGTGATCCTTCATCTCATGCTT 1138  
DB 531 KYSYSAARKARCWYRGYWGMMWKRYKRYMYWYKMMWYKRYKSKYKMSYVASC 590  
QY 1139 ATATGTG-GAGTAGATAGATTCTCACCATAATTAGATGACAAAGCAGAGATTGTGT 1197  
DB 591 MKSARKAGAKCKRSKMSAWSKMSRSCRKCKKASRKSARAKRYAMMGMTSGSRMSRWKS 650  
QY 1198 TTTATCTGTTGGTAAATACGTTTCTCCAGTCTGTATATAAGACCTCCACAGATATAAA 1257  
DB 651 YTCWRKWSKSTCTCTWYMSKYTYAKYSGYWRYYRAMCWYMYRYRYRSYMYMA 710  
QY 1258 GTCCTATGCAACAAGAAATGTCAATACATCTCTTAGTCTCATATTATTTTCAATAG 1317  
DB 711 WYTSSTRMAMTGMKYSGRYTSWYKCKSKYKRSWYMSWYMSWAKTWKMRRYATRMW 770  
QY 1318 ATACCGGTTTTTACTACAACTCAATAAGATGACAGATGAATGGTTAGTGACTGT 1377  
DB 771 MWYRYSKMYTWCTMGWYMYWYMYRMYKMYKCTKYWYWSATYWTGTWAAWMAKT 830  
QY 1378 TTATAAAGAGATATAAAGATACTATCATCTTTGAGGCAATAAGGGAGGAGATT 1437  
DB 831 KMRMGMTGAKTRGRARARVWYKWTWCATKRMWTKGKAKWATWYMAKWRKYYSWMR 890  
QY 1438 CAGCAAAACAGTGTCTTACAAGTGGAAAAACAAGTAAAGTGAAGTCCCTCTCTGA 1497

DB 891 AWYVYKTRTRYKTCWKARWGSWAYRWMWKGSAKQWMMWKGGRWGTKYWYWYCTTW 950  
QY 1498 CAGATCAATGCCACAGCTTGAGCTTTAGCCAGCCACATCATCATGTAATGCTTTCTCTG 1557  
DB 951 KMACGRATKYMCCAGWMMYSYSWTRYWRTWRWMMASSTAKRMARMWMTKRAWKS 1010  
QY 1558 ATAAGCCTGTTCATAAATCTCTTTGCAAAAGCTCTGCTACTTACCA 1603  
DB 1011 ARAYWKGAGCCTACACTACACAGAAATGTTGTTTATTTCCA 1056  
  
RESULT 8  
ABX40871/c  
ID ABX40871 standard; cDNA; 337 BP.  
XX  
AC ABX40871;  
XX  
DT 20-FEB-2003 (first entry)  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #6036.  
XX  
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.  
XX  
OS Bos Taurus.  
XX  
PN US2002137139-A1.  
XX  
PD 26-SEP-2002.  
XX  
PF 24-SEP-2001; 2001US-00960352.  
XX  
PR 12-JAN-1999; 99US-0115707P.  
XX  
PR 11-JAN-2000; 2000US-00480902.  
XX  
PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
XX  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX  
DR MPI; 2003-110599/10.  
XX  
PT New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and analysis,  
PT cattle breeding, or for genetically improving cattle.  
XX  
PS Claim 2; SEQ ID NO 6036; 245pp; English.  
XX  
CC The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived from  
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second  
CC nucleic acid molecule comprising any of 1512 nucleotide sequences,  
CC appearing as ABX34836-ABX49947, or complements of them. Also included are  
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic  
CC acid linked to a promoter and a 3' non-translated sequence that  
CC functions in the cell to cause termination of transcription and addition  
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
CC (2) determining a level or pattern of a molecule in a bovine cell or  
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
CC of the 1512 nucleic acid sequences or its complement or fragment) with a  
CC complementary nucleic acid molecule obtained from the bovine cell or  
CC tissue, where hybridisation between the marker nucleic acid and the  
CC complementary nucleic acid permits the detection of the molecule; and (b)  
CC detecting the level or pattern of the complementary nucleic acid, where  
CC the detection of the complementary nucleic acid is predictive of the  
CC level or pattern of the molecule. The LMFD nucleic acid is used for  
CC determining a level or pattern of a molecule in a bovine cell or tissue.  
CC It is useful for genome mapping, gene identification and analysis, cattle  
CC breeding, preparation of constructs for use in cattle gene expression, or

CC for genetically improving cattle. The present sequence is one of the  
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The  
CC present sequence was not shown in the specification but was obtained in  
CC electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docID=20020137139  
XX

SQ Sequence 337 BP; 150 A; 35 C; 40 G; 112 T; 0 U; 0 Other;  
Query Match 2.0%; Score 47.2; DB 8; Length 337;  
Best Local Similarity 53.9%; Pred. No. 0.39; Mismatches 0; Indels 0; Gaps 0;  
Matches 97; Conservative 0;  
QY 2185 AAGAGTAAGACTCTGAGTCAGAAATTCGGCGCTAAGCTCCCTCAACTACAGAAAAGTCAC 2244  
Db 289 AATAAATATATATATTTTACAAATTTAAATTTAAGTTACATTATATAAAAAAATGTGAT 230  
QY 2245 AATAAAAAATCAACATGATGTTCTATTTTGTGTTTCTCTGCTGTGATGTTAAATGATAT 2304  
Db 229 ATTTAATATATTTAAATATTTTTTTTTTTTATTTTATTTTTTTTTTTTTTTTTTTTTTTT 170  
QY 2305 TATTATTTTTTTTTTAGGCGTGAATCTGATGAAAGAAAGTTAGGCGCCACGATAATC 2364  
Db 169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTAAGTGTATTACTTTTCAATAC 110

RESULT 9  
ABL15880  
ID ABL15880 standard; cDNA; 2636 BP.  
XX  
AC ABL15880;  
XX  
XX  
26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42122.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
XX  
PD 27-SRP-2001.  
XX  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX  
WIPI; 2001-656860/75.  
DR P-PSDB; ABB71777.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX  
PS Claim 1; SEQ ID NO 42122; 21pp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 2636 BP; 849 A; 414 C; 497 G; 876 T; 0 U; 0 Other;  
Query Match 2.0%; Score 47; DB 4; Length 2636;  
Best Local Similarity 52.3%; Pred. No. 0.86;  
Matches 104; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 1821 ATTGCCTGCACCTATTTTGAAGAACTCTATAGAGGGGAATACAGAGGAAGAAAAACATT 1880  
Db 540 ATTGACTTCGAATATCTTTAAATATACCAGCCCTGTCTGGTTCTTCGAATTAACAAT 599  
QY 1881 CTGATTTTATTTGCGATTGCGATAATCTTATGCAATTAGCTAATTCAGTAGGCGATTCC 1940  
Db 600 ATAAATTTATTTTAAATGGTGAATTAATTCGTTCTTAAAAAATTCAGGTATATTTTATTC 659  
QY 1941 ACAGAGAAATTTAAATAGAAATATATGTAAGGAATATTTTATGATAAGACTGTTTGAAGA 2000  
Db 660 TTAGGGAATTAAGTATTTATTTGCTGCTAGTTTATTTATTTTAAAAAAGGTATAGACAA 719  
QY 2001 ATTACACAGAGAGGAAAT 2019  
Db 720 ATTAGTAGATGTGTAATTT 738  
RESULT 10  
ADD29587  
ID ADD29587 standard; mRNA; 662 BP.  
XX  
AC ADD29587;  
XX  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Mouse tumour suppressor mRNA SEQ ID NO:36.  
XX  
KW ss; mouse; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
XX  
OS Mus musculus.  
XX  
XX  
PN WO2003058201-A2.  
XX  
XX  
PD 17-JUL-2003.  
XX  
XX  
PF 31-DEC-2002; 2002WO-US041825.  
XX  
XX  
PR 31-DEC-2001; 2001US-0345317P.  
XX  
XX  
PA (QUAR-) QUARK BIOTECH INC.  
PA (CLEV-) CLEVELAND CLINIC FOUND.  
XX  
XX  
PI Feinstein E, Gudkov AV;  
XX  
XX  
WIPI; 2003-598393/56.  
XX  
XX  
PT Diagnosing cancer comprises determining the polypeptide or polynucleotide  
PT levels e.g., hepatic lipase, in a sample from a subject, where a higher  
PT level compared to that in a subject free of cancer is indicative of  
PT cancer.  
XX  
XX  
PS Disclosure; SEQ ID NO 36; 272pp; English.  
XX  
XX  
CC The invention relates to a novel method for diagnosing a cancer in a  
CC subject, the method comprises determining, in a sample from the subject,  
CC the level of at least one polypeptide, where a higher level of the  
CC polypeptide compared to the level of the polypeptide in a subject free of  
CC cancer is indicative of cancer. The polypeptide is selected from any of  
CC the polypeptides encoded by the polynucleotides listed in the  
CC specification and polypeptides which are at least 70% homologous to the  
CC polypeptides. The method of the invention has cytostatic activity, and  
CC may have a use in gene therapy. The method is useful in identifying  
CC markers specific for one or several types of cancer, depending on the  
CC tissue origin, which may be used in numerous diagnostic and prognostic  
CC applications as well as cancer type-specific targets for therapeutic  
CC intervention. The compounds that modulate the activity of a tumour  
CC suppressor gene are useful in the treatment of cancer or as anti-cancer  
CC









**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 08:01:56 ; Search time 221.713 Seconds  
(without alignments)  
7633.255 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agctcttcgagcagaagg.....atctgaagctcactattcag 2381

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues  
Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	64.2	2.7	1141	4	US-09-806-708B-22
C 2	57.8	2.4	1141	4	US-09-806-708B-22
C 3	48.4	2.0	832	4	US-09-621-976-2813
C 4	44.2	1.9	515	3	US-09-385-982-474
C 5	43.8	1.8	7218	1	US-08-232-463-14
C 6	43	1.8	832	4	US-09-621-976-2813
C 7	42.6	1.8	611	3	US-09-385-982-376
C 8	42.4	1.8	452	4	US-09-401-064-138
C 9	42.2	1.8	652	4	US-09-270-767-10640
C 10	42.2	1.8	1257	4	US-09-270-767-15058
C 11	41.6	1.7	1273	4	US-09-270-767-14731
C 12	41.2	1.7	196	4	US-09-442-054A-42
C 13	41.2	1.7	196	4	US-09-442-054A-42
C 14	40.4	1.7	640681	4	US-09-790-988-1
C 15	40.2	1.7	1055	4	US-09-806-708B-23
C 16	40	1.7	733	3	US-09-270-767-10136
C 17	39.6	1.7	43360	3	US-09-453-702B-206
C 18	39.6	1.7	45325	3	US-09-453-702B-261
C 19	39.2	1.6	4860	4	US-09-270-767-10504
C 20	38.6	1.6	3687	4	US-09-891-641-1
C 21	38.4	1.6	6656	4	US-10-204-708-75
C 22	38.4	1.6	9573	4	US-09-220-132-168
C 23	38	1.6	55827	4	US-09-813-133A-3
C 24	38	1.6	640681	4	US-09-790-988-1
C 25	37.6	1.6	495	4	US-09-270-767-4884
C 26	37.6	1.6	495	4	US-09-270-767-20166
C 27	37.6	1.6	19124	2	US-08-487-826B-13

28	37.4	1.6	396	4	US-09-640-173-33	Sequence 33, Appl
29	37.4	1.6	396	4	US-09-713-550-33	Sequence 33, Appl
30	37.4	1.6	396	4	US-09-825-294-33	Sequence 33, Appl
31	37.4	1.6	396	4	US-09-970-966-33	Sequence 33, Appl
C 32	37.4	1.6	663	3	US-08-998-416-187	Sequence 187, App
C 33	37.4	1.6	696	3	US-08-998-416-179	Sequence 779, App
C 34	37.4	1.6	719	3	US-08-998-416-1138	Sequence 1138, Ap
C 35	37.4	1.6	856	3	US-08-998-416-289	Sequence 289, App
C 36	37.2	1.6	1046	1	US-08-361-467B-4	Sequence 4, Appli
C 37	37.2	1.6	1046	1	US-08-484-332C-4	Sequence 4, Appli
C 38	36.8	1.5	659	4	US-09-270-767-29436	Sequence 29436, A
C 39	36.8	1.5	1643	4	US-09-270-767-13457	Sequence 13457, A
C 40	36.6	1.5	1227	4	US-09-543-681A-2270	Sequence 2270, Ap
C 41	36.6	1.5	1262	3	US-09-227-357-27	Sequence 27, Appl
C 42	36.6	1.5	3198	3	US-08-942-306B-48	Sequence 48, Appl
C 43	36.6	1.5	3198	3	US-08-938-373B-48	Sequence 48, Appl
C 44	36.6	1.5	3199	4	US-09-945-249-10	Sequence 10, Appl
C 45	36.6	1.5	3199	4	US-09-041-990-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-09-806-708B-22/c  
; Sequence 22, Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT APPLICATION NUMBER: US/09/806,708B  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 60/147,133  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; TYPE: DNA  
; LENGTH: 1141  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1141)  
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters  
US-09-806-708B-22

Query Match	2.7%	Score 64.2	DB 4	Length 1141
Best Local Similarity	12.7%	Pred. No. 2.3e-07		
Matches	136	Conservative 379	Mismatches 550	Indels 10
Gaps	3			
Qy	830	TTTTCTGCATTATCCTTCAACATTTAAACCTGGGATCTATGGAATCAAAACAGTTGGG	889	
Db	1129	TTTKTKYKKNNNNNNGKDNRMDATKWSATGTAWWTHAKRGATMCWYWTGTN	1070	
Qy	890	TAAATTCATCTAGCAGCATCACTACTGTAGGATGACAGACAGAGCATTCCT	949	
Db	1069	RRWCMTYAMRTWYTRSNANWSCATKBMWMTKMYATKYRYTAWYAMWCANNNNNWCATN	1010	
Qy	950	GAATGGGCTATAATAGAGAATACCTAGAGGTGCTCGAATTTAGACTACCTATTAAA	1009	
Db	1009	GYAKSATNNMAYATRWAAAYAAAKWAGNNMMYGAAGKWKGMATMGWBA	950	
Qy	1010	GAGTGAGGACAGAAATGGAGAAATATCATCGCAATTTCTGTAGCTACGACTAGACTGAA	1069	
Db	949	DTAGKCMNNNNNTTIDVRRMAMKAKNNNNNNYATACYNRAATNNKMAATMMKWTGHAH	890	
Qy	1070	GGTTTCGAACTGAACCGAG-----TTTCCAAACTACCTGTGATGTTTCAGTGGAT	1122	
Db	889	SKRTRHHTTCRTRKYNNNNNNNARTVYTHHAARRMMNAWTRTNNNNNNNNNACRNT	830	
Qy	1123	CCVTCATCTCATGCTTATTATGTGGAGTAGAATCTCCACCAATTAGAATGGACAA	1182	



Matches	55;	Conservative	0;	Mismatches	18;	Indels	0;	Gaps	0;
QY	1623	AAAGATGGCATTAAACGGTACTTGGAAAAATAGAGAAAAATGAGAACTATATGAAAAATTCAT	1682						
Db	23	AATCATGGCGTTTGACAGCACTTGGAGGTTAGACCGGAGTGAACCACTATGACCAAGTTTCAT	82						
QY	1683	GGAAGCAATGGGT	1695						
Db	83	GGAAAAAATGGGT	95						
<p>RESULT 5</p> <p>US-08-232-463-14/C</p> <p>Sequence 14, Application US/08232463</p> <p>Patent No. 5670367</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: DORNER, F.</p> <p>APPLICANT: SCHEIFLINGER, F.</p> <p>APPLICANT: FALKNER, F. G.</p> <p>TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS</p> <p>NUMBER OF SEQUENCES: 52</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Foley &amp; Lardner</p> <p>STREET: 1800 Diagonal Road, Suite 500</p> <p>CITY: Alexandria</p> <p>STATE: VA</p> <p>COUNTRY: USA</p> <p>ZIP: 22313-0299</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent In Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/232,463</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 435</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/07/935,313</p> <p>FILING DATE:</p> <p>APPLICATION NUMBER: EP 91 114 300.6</p> <p>FILING DATE: 26-AUG-1991</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: BENT, Stephen A.</p> <p>REGISTRATION NUMBER: 29,768</p> <p>REFERENCE/DOCKET NUMBER: 30472/114 IMMU</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (703)836-9300</p> <p>TELEFAX: (703)683-4109</p> <p>TELEX: 899149</p> <p>INFORMATION FOR SEQ ID NO: 14:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 7218 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>IMMEDIATE SOURCE:</p> <p>CLONE: pTZgpt-F1s</p> <p>US-08-232-463-14</p>									
<p>Query Match 1.8%; Score 43.8; DB 1; Length 7218;</p> <p>Best Local Similarity 8.4%; Pred. No. 0.18;</p> <p>Matches 36; Conservative 204; Mismatches 191; Indels 0; Gaps 0;</p>									
QY	628	TATCTGTGTGATAAGCAACATTCATTTTCAGCATTCAGGATTTTACATTTTGAAGCTAATA	687						
Db	1474	TATCTATGCACTAGTTAAAGAGATAGAAAGATTGGTACRRRRRRRRRRRRRRRRRRRRRR	1415						
QY	688	GACAGCAGATTTGGTGCGCTCCATAGGAACAGACTAATACTAATCCTGAGTTAGTACAA	747						
Db	1414	RR	1355						
QY	748	GCAGATTTAGCACCAGCAAAATTTGCTCAGTTTCAAGTAGCAGCTATCTTTGTGGGAAGAAG	807						



```
; NAME/KEY: misc feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-401-064-138

Query Match      1.8%; Score 42.4; DB 4; Length 452;
Best Local Similarity 53.7%; Pred. No. 0.11;
Matches 88; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1790 TCCCTTGTGCTGCTATTTTGGCCCTGCACATTGCCCTGCACATTTTGAAGAAGACTC 1849
Db 379 TCCCTTGTGCTGCTATTTTGGCCCTGCACATTGCCCTGCACATTTTGAAGAAGACTC 320
Qy 1850 TATAGAGGGGAATCAAGGAAGAAAAACATTCCTGATTTTATTTGCAITGGCATATCTTA 1909
Db 319 TATAGAGCTAAAAAATTTGAACAACAACAAAAAATGCTGTGATTTTCCAAATGTAA 260
Qy 1910 TGCATTTAGCTAAATCCAGTAGAGGCATTCAGCAGAGAAATTTAA 1953
Db 259 TACATTTACTTTAGCAITTTGAAGCCACTTTGAACCTCGAGATAA 216

RESULT 9
US-09-270-767-10640/c
; Sequence 10640, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10640
; LENGTH: 652
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-10640

Query Match      1.8%; Score 42.2; DB 4; Length 652;
Best Local Similarity 50.8%; Pred. No. 0.15;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 1821 ATTCCTCCCTGCACTTATTTTGAAGAGCTCTATAGAGGGAATCAAGGAAGAAAAACATT 1880
Db 355 ATTGAATTCGAATTAATCTTAAATACAGCCCTGTTCTGGTTCTTCGAATTAACAAT 296
Qy 1881 CTGATTTTATTCGCAATTCGATAATCTTATGCAATTTAGCTAAATCCAGTAGAGGCATTC 1940
Db 295 TGAATATTTTAAATGTCAAATTAATTTGTTCTTAAAAAATTCAGGTATATTTATTC 236
Qy 1941 AGCAGAAATTAATAGAAATTAATATGTAAGGAATATTTATTTGATAGACTGTTTGAAA 2000
Db 235 TTAGGGAATTAAGATTAATTTGCTGCTAGTTTATTTATGTTAAAAAAGGTATAGACAA 176
Qy 2001 ATTACACAGAGGAGAAATTT 2019
Db 175 ATTAGTAGATGTGTAATTT 157

RESULT 10
US-09-270-767-15058/c
; Sequence 15058, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15058
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15058

Query Match      1.8%; Score 42.2; DB 4; Length 1257;
Best Local Similarity 53.3%; Pred. No. 0.2;
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 1187 GAGATTGTGTTTATCTGTTGGTAAATAGCTTTTCCAGTTGTATAAAGACCTCC 1246
Db 364 GAGTTTATCATTTATATTTTAAAGTTTCAACTGCAACACCCAGCAGCTACTC 305
Qy 1247 ACCAGTATAAGTCCCTATGCAACAAGAAAAATGTCATACATCTCTTAGTCTCATTA 1306
Db 304 ACAGCAAAACGTTACAGGAGGAAGAAAGAAAGATTAAGAGTGGTATTTCTTCACAA 245
Qy 1307 ATTTTCATTAGTAGCGCGTTTTTTTACTACAACCTCAATTAAGATGAA 1353
Db 244 TGTTTTATGCAATAAAAGGTGTGCCATTTCATATCAATATAAAGTA 198

RESULT 11
US-09-270-767-14731/c
; Sequence 14731, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14731
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14731

Query Match      1.7%; Score 41.6; DB 4; Length 1273;
Best Local Similarity 62.5%; Pred. No. 0.3;
Matches 65; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 2264 TGTTCTATTTTGTGTTTCTCTGCTTGATGTTAATTAATTAATTTATTTTATTTAGG 2323
Db 1188 TTTTCTATTTTGTGTTTCTCTGCTTGATGTTAATTAATTAATTTATTTTATTTTTCG 1129
Qy 2324 CGTGAATGTGATGAAAGAAAGTTAGGAGCCAGCAATAAATCTGA 2367
Db 1128 TTGATTTGAATTAAGCGAACTTTATATCCCATCTAAATCCGA 1085

RESULT 12
US-09-442-054A-42
; Sequence 42, Application US/09442054A
; Patent No. 6770738
; GENERAL INFORMATION:
; APPLICANT: Ecker, David J.
; APPLICANT: Buchardt, Ole
; APPLICANT: Egholm, Michael
; APPLICANT: Berg, Rolf H.
; APPLICANT: Mollegaard, Neils E.
; TITLE OF INVENTION: Higher Order Structure And Binding Of Peptide Nucleic Acids
; FILE REFERENCE: ISIS4290
; CURRENT APPLICATION NUMBER: US/09/442,054A
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 08/471,907
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/054,363
; PRIOR FILING DATE: 1993-04-26
; PRIOR APPLICATION NUMBER: PCT/ EP92/01219
```

; PRIOR FILING DATE: 1992-05-19  
; NUMBER OF SEQ ID NOS: 89  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42  
; LENGTH: 196  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: No. 6770738el Sequence  
US-09-442-054A-42

Query Match 1.7%; Score 41.2; DB 4; Length 196;  
Best Local Similarity 64.9%; Pred. No. 0.15;  
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2227 AACTACAGAAAAAGTCACATAAAAAATGCAAAACATGATGTTCTATTGTTTCTCTGC 2286  
Db 62 AA 121  
QY 2287 TTGATGTTAAATTGATTATTATTATTTT 2320  
Db 122 TTTT 155

## RESULT 13

US-09-442-054A-42/c  
; Sequence 42, Application US/09442054A  
; Patent No. 6770738

; GENERAL INFORMATION:  
; APPLICANT: Ecker, David J.

; APPLICANT: Buchardt, Ole

; APPLICANT: Egholm, Michael

; APPLICANT: Berg, Rolf H.

; APPLICANT: Mollegaard, Neils E.

; TITLE OF INVENTION: Higher Order Structure And Binding Of Peptide Nucleic Acids

; FILE REFERENCE: ISIS4290

; CURRENT APPLICATION NUMBER: US/09/442,054A

; PRIOR FILING DATE: 2002-05-07

; PRIOR APPLICATION NUMBER: 08/471,907

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: 08/054,363

; PRIOR FILING DATE: 1993-04-26

; PRIOR APPLICATION NUMBER: PCT/ EP92/01219

; PRIOR FILING DATE: 1992-05-19

; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 42

; LENGTH: 196

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: No. 6770738el Sequence

US-09-442-054A-42

Query Match 1.7%; Score 41.2; DB 4; Length 196;  
Best Local Similarity 64.9%; Pred. No. 0.15;  
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2227 AACTACAGAAAAAGTCACATAAAAAATGCAAAACATGATGTTCTATTGTTTCTCTGC 2286  
Db 135 AA 76  
QY 2287 TTGATGTTAAATTGATTATTATTATTTT 2320  
Db 75 TTTT 42

## RESULT 14

US-09-790-988-1/c

; Sequence 1, Application US/09790988

; Patent No. 6632935

; GENERAL INFORMATION:

; APPLICANT: SHIGENOBU, SHUJI

; APPLICANT: WATANABE, HIDEKI  
; APPLICANT: HATTORI, MASAHIRA  
; APPLICANT: SAKAKI, YOSHIYUKI  
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
; FILE REFERENCE: 081356/0159  
; CURRENT APPLICATION NUMBER: US/09/790,988  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: JP2000-107160  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 1.7%; Score 40.4; DB 4; Length 640681;  
Best Local Similarity 51.7%; Pred. No. 14;  
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 308 TGTGTTGGTGCACAACTACAGCAACTTTGTTGCAACAATAACAATCTAAGTTGTTTCTT 367  
Db 501857 TCTGTAGTTTATTATTAAACACATATATATCACTAATAAAGTTGTTGTTTTTATT 501798  
QY 368 TTCCTCTTTCCTTAACCTCTGTACAGTCTAAAGGTGAAGAGTAGCTATTGAGTTACTTC 427  
Db 501797 TATTTTTTACACAATTTATATTGTAAGTTTTTCAATACATACTAATAGATTAAAGCT 501738  
QY 428 CTCGTCATCCTCTAGCCAGATTAGCATGATTCCAAAATGACCTGAGTGAATGGA 485  
Db 501737 ATATCCATCATTTGGTGAATAACAATTTTCTTAGAATAGAACCTGATTCTTATGGA 501680

## RESULT 15

US-09-806-708B-23/c

; Sequence 23, Application US/09806708B

; Patent No. 6784342

; GENERAL INFORMATION:

; APPLICANT: The University of British Columbia

; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants

; FILE REFERENCE: 4810-58741

; CURRENT APPLICATION NUMBER: US/09/806,708B

; CURRENT FILING DATE: 2001-04-03

; PRIOR APPLICATION NUMBER: US 60/147,133

; PRIOR FILING DATE: 1999-08-04

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 23

; LENGTH: 1055

; TYPE: DNA

; ORGANISM: Artificial sequence

; FEATURE:

; NAME/KEY: promoter

; LOCATION: (1)..(1055)

; OTHER INFORMATION: consensus sequence of A.t. and L.a. FAEI promoters

US-09-806-708B-23

Query Match 1.7%; Score 40.2; DB 4; Length 1055;  
Best Local Similarity 23.3%; Pred. No. 0.64;  
Matches 140; Conservative 119; Mismatches 340; Indels 3; Gaps 2;

QY 271 AAAACAATTAATTTCTGGTGACAGTCAGTAGAGCCCTGTTTGGGTGCAACTACAGCA 330  
Db 790 AAATANAATRTNN 731  
QY 331 CTTTGTGCAACAATAACAATCTAAGTTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 390  
Db 730 NAYKCAANNAAAAAAGTCNNNTAATTTTCTCTTAANTATYTRATT-AKTTTTEWA 672  
QY 391 CAGTCTAAAGGTGAAGTAGTAGCTATTGAGTTACTTCCCTCTGCACTCTCTTAGCAGAT 450  
Db 671 TTSTKMTNNNNNAWACTNN 612





**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 07:33:41 ; Search time 8247.18 Seconds  
(without alignments)  
10520.322 Million cell updates/sec

Title: US-10-099-663-1  
Perfect score: 2381  
Sequence: 1 agcttctgcgcgagaaag.....atctgaagtcactattcag 2381

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_btc.\*
- 4: gb\_est3.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_gss1.\*
- 9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	752.6	31.6	1473	8	CC250935	CH261-17C
2	99	4.2	598	6	CF250818	esa017_e0
3	95	4.0	595	6	CF250947	esa019_a0
4	85	3.6	705	5	B0355106	B0355106 603474288
5	84	3.5	759	5	B0265304	B0265304 603508640
6	81	3.4	829	5	B0296697	B0296697 603741965
7	79.4	3.3	885	5	B0123280	B0123280 603003274
8	79	3.3	814	5	B0234099	B0234099 603792609
9	78.8	3.3	1080	5	B0123336	B0123336 603149475
10	75.8	3.2	426	6	CD739695	CD739695 4028031_1
11	74	3.1	576	6	CD739141	CD739141 4026481_1
12	70.4	3.0	480	6	CD733346	CD733346 4045132_1
c 13	62.4	2.6	1101	9	CNS0039G	AL063921 Drosophil
14	61.4	2.6	885	5	B0123171	B0123171 603147391
15	60.4	2.5	443	5	BX276255	BX276255 BX276255
16	57.2	2.4	683	7	CN793516	CN793516 4128553_B
17	55.8	2.3	657	7	CK945920	CK945920 4070423_B
18	54.2	2.3	572	7	CK834687	CK834687 4059069_B
19	54.2	2.3	673	7	CK947902	CK947902 4072726_B
20	54.2	2.3	686	7	CK957350	CK957350 4097834_B
21	54.2	2.3	708	7	CK960577	CK960577 401917_B
22	54	2.3	481	4	BM430667	BM430667 lDuo03All
23	54	2.3	489	6	CB224035	CB224035 lJEB30F5
24	54	2.3	497	4	BM430956	BM430956 lDuo07F03

25	54	2.3	548	4	BM432416	BM432416 lJEB74B4.a
26	54	2.3	568	4	BM431104	BM431104 lDuo09G09
27	54	2.3	671	7	CK980160	CK980160 4111791_B
28	54	2.3	679	7	CK947198	CK947198 4071820_B
29	54	2.3	682	7	CK948900	CK948900 4074042_B
c 30	54	2.3	684	7	CK947048	CK947048 4071436_B
31	54	2.3	702	7	CK944155	CK944155 4068182_B
32	54	2.3	720	7	CK946512	CK946512 4071131_B
33	53	2.2	372	4	BM430450	BM430450 lDuo34E2.
34	53	2.2	414	4	BM433066	BM433066 lJEB9E2.a
35	53	2.2	416	4	BM432621	BM432621 lJEB11H7.
36	53	2.2	546	7	CN098758	CN098758 EC2CAA15A
c 37	53	2.2	576	7	CN098757	CN098757 EC2CAA15A
38	52.6	2.2	660	7	CK954372	CK954372 4094190_B
39	52.6	2.2	1101	9	CNS0039G	AL063921 Drosophil
40	52.4	2.2	503	4	BM432863	BM432863 lJEB7B11
41	52	2.2	339	4	BM430492	BM430492 lDuo35B12
42	52	2.2	371	4	BM432453	BM432453 lJEB74F5.a
43	52	2.2	371	4	BM433013	BM433013 lJEB7B11
44	52	2.2	415	4	BM430284	BM430284 lDuo32C9.
45	52	2.2	433	4	BM431869	BM431869 lJEB13B10

ALIGNMENTS

RESULT 1  
CC250935/c  
LOCUS CH261-17C7.Sp6.1 CH261 Gallus gallus genomic clone CH261-17C7, linear GSS 13-MAY-2003  
DEFINITION genomic survey sequence.  
ACCESSION CC250935  
VERSION CC250935.1 GI:30587685  
KEYWORDS GSS.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 1473)  
AUTHORS Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.  
TITLE Gallus gallus BAC End Reads  
JOURNAL Unpublished (2003)  
COMMENT Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: submissions@watson.wustl.edu  
Insert Length: 182000 Std Error: 0.00  
Seq primer: Sp6 ATTTAGTGACACTATAG  
Class: BAC ends  
High quality sequence start: 239  
High quality sequence stop: 912.  
Location/Qualifiers  
source  
1..1473  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/strain="Red Jungle Fowl"  
/db\_xref="taxon:9031"  
/clone="CH261-17C7"  
/sex="female"  
/cell\_line="UCD001, inbred 256"  
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: http://www.chori.org/bacpac"

ORIGIN  
Query Match 31.6%; Score 752.6; DB 8; Length 1473;  
Best Local Similarity 98.3%; Pred. No. 4.1e-168;  
Matches 793; Conservative 0; Mismatches 9; Indels 5; Gaps 3;  
QY 1580 TTTGCAAGCTCTGCTACTTACCAGAAGTC-TGCTACAGACAGAAAGATGGCATTTAA- 1637

Db 1042 TTGCAAAAGTTCGCTTCTTACCAGAAATCTCCCTACACAGAGATGGCATTTAAC 983  
Qy 1638 CGGTACTTGGAATAAGAGAAAATAGAACTATGAAAAATTCATGGAAGCAATGGGTAA 1697  
Db 982 CGGTACTTGGAATAAGAGAAAATAGAACTATGAAAAATTCATGGAAGCAATGGGTAA 923  
Qy 1698 GCCTTACTTTTTGAATGCCTTCTAAAAGCAGATACCACTACGGCGGATACAAACTTA 1757  
Db 922 GCCTTACTTTTTGAATGCCTTCTAAAAGCAGATACCACTACGGCGGATACAAACTTA 863  
Qy 1758 AGCTGTTTCATGAACCTACCACTCGCTTAACCTGCTTCTTGTGTGCTGCTATTTTGCCTTGG 1817  
Db 862 AGCTGTTTCATGAACCTACCACTCGCTTAACCTGCTTCTTGTGTGCTGCTATTTTGCCTTGG 803  
Qy 1818 CACATTCGCCCTGCACATTTTGAAGAAGACTCTATAGAGGGGAATACAGAGAAAGAAAAC 1877  
Db 802 CACATTCGCCCTGCACATTTTGAAGAAGACTCTATAGAGGGGAATACAGAGAAAGAAAAC 743  
Qy 1878 ATTCTGATTTTATTTGCAATTCGATATCTTATGCAATTTAGCTAATTCAGTAGAGGCAT 1937  
Db 742 ATTCTGATTTTATTTGCAATTCGATATCTTATGCAATTTAGCTAATTCAGTAGAGGCAT 683  
Qy 1938 TCCAGCAGAAAATTAATAGAAATATATGTAAGGAATATTTATTTTGATAGACTGTTTGA 1997  
Db 682 TCCAGCAGAAAATTAATAGAAATATATGTAAGGAATATTTATTTTGATAGACTGTTTGA 623  
Qy 1998 AAAATTACAGAGGGAATTCGTGCTCCAGTTTTCAGAGAACACACATGATTTGAGT 2057  
Db 622 AAAATTACAGAGGGAATTCGTGCTCCAGTTTTCAGAGAACACACATGATTTGAGT 563  
Qy 2058 CATTTTAAACATGCTAGTGCCTTACTTTAAAGCTTCACAACTGCTGTAATATGATGATA 2117  
Db 562 CATTTTAAACATGCTAGTGCCTTACTTTAAAGCTTCACAACTGCTGTAATATGATGATA 503  
Qy 2118 ACATACTATCCTAGTTGGATAGTGTGTTGTTATTCAGAGCTGAACTGCTCAGTGAAA 2177  
Db 502 ACATACTATCCTAGTTGGATAGTGTGTTGTTATTCAGAGCTGAACTGCTCAGTGAAA 443  
Qy 2178 GGTGGAGAGATGAAGCTCTGATCAGATTCCTGGCTAAGCTCCCTCACTACAGAAA 2237  
Db 442 GGTGGAGAGATGAAGCTCTGATCAGATTCCTGGCTAAGCTCCCTCACTACAGAAA 383  
Qy 2238 AAGTCACATATAAAATGCAAACTATGTTCTATTTTGTCTGCTGTTGATTTAT 2297  
Db 382 AAGTCACATATAAAATGCAAACTATGTTCTATTTTGTCTGCTGTTGATTTAT 323  
Qy 2298 TGATTATTTATTA---TTTTTTTTTTAGGCGTGAATGTGATGAAAAGAAAGTTAGAGCC 2354  
Db 322 TGATTATTTATTTATTTTTTTTTTTTTAGGCGTGAATGTGATGAAAAGAAAGTTAGAGCC 263  
Qy 2355 CACGATAATCTGAAGCTCACTATTTCAG 2381  
Db 262 CACGATAATCTGAAGCTCACTATTTCAG 236

RESULT 2  
CF250818  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

CF250818 598 bp mRNA linear EST 07-AUG-2003  
esa017\_s05 Eimeiria tenella-infected caecal tonsil Gallus gallus  
cDNA, mRNA sequence.  
CF250818  
CF250818.1 GI:33484073  
EST.  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 598)  
Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,  
Chausee,A.M. and Zoorob,R.  
A collection of chicken ESTs from activated immune cells

JOURNAL  
COMMENT

Unpublished (2003)  
Contact: Zoorob R  
UPR 1983  
CNRS  
7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@vjf.cnrs.fr.  
Location/Qualifiers  
1..598  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone\_lib="Eimeiria tenella-infected caecal tonsil"  
/note="Organ: Caecal tonsil; Vector: pTriplex2"

FEATURES  
source

1..598  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone\_lib="Eimeiria tenella-infected caecal tonsil"  
/note="Organ: Caecal tonsil; Vector: pTriplex2"

ORIGIN

Query Match 4.2%; Score 99; DB 6; Length 598;  
Best Local Similarity 100.0%; Pred. No. 1.8e-12; Mismatches 0; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1596 ATTACCAGAAAGTCTGCCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGA 1655  
Db 3 ACTTACCAGAAAGTCTGCCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGA 62  
Qy 1656 GAAAATGAGAACTATGAAAAATTCATGGAAGCAATGGG 1694  
Db 63 GAAAATGAGAACTATGAAAAATTCATGGAAGCAATGGG 101

RESULT 3  
CF250947  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CF250947 595 bp mRNA linear EST 07-AUG-2003  
esa019\_a07 Eimeiria tenella-infected caecal tonsil Gallus gallus  
cDNA, mRNA sequence.  
CF250947  
CF250947.1 GI:33484202  
EST.  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 595)  
Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,  
Chausee,A.M. and Zoorob,R.  
A collection of chicken ESTs from activated immune cells  
Unpublished (2003)  
Contact: Zoorob R  
UPR 1983  
CNRS  
7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@vjf.cnrs.fr.  
Location/Qualifiers  
1..595  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone\_lib="Eimeiria tenella-infected caecal tonsil"  
/note="Organ: Caecal tonsil; Vector: pTriplex2"

FEATURES  
source

1..595  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone\_lib="Eimeiria tenella-infected caecal tonsil"  
/note="Organ: Caecal tonsil; Vector: pTriplex2"

ORIGIN

Query Match 4.0%; Score 95; DB 6; Length 595;  
Best Local Similarity 100.0%; Pred. No. 1.6e-11; Mismatches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1600 ACCAGAAGTCTGCCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAA 1659  
Db 3 ACCAGAAGTCTGCCCTACAGACAGAAAGATGGCATTTAACGGTACTTGGAAAATAGAGAA 62  
Qy 1660 AATGAGAACTATGAAAAATTCATGGAAGCAATGGG 1694

```

Db      63 AATGAGAACTATGAAAAATTCATGGAAGCAATGGG 97
|||||
RESULT 4
BU355106
LOCUS   603474288F1 CSEQCHN70 Gallus gallus CDNA clone CHEST355022 5', mRNA
DEFINITION
ACCESSION
BU355106
VERSION 603474288F1 GI:25863107
KEYWORDS
SOURCE   Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 705)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE   A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..705
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST355022"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN70"
/notes="Organ: hearts; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
ORIGIN
Query Match 3.6%; Score 85; DB 5; Length 705;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1610 TGCTACAGACAGAAAGATGCGATTTAACGCTACTTCGAAATAGAGAAAAATGAGAACT 1669
|||||
Db 1 TGCTACAGACAGAAAGATGCGATTTAACGCTACTTCGAAATAGAGAAAAATGAGAACT 60
|||||
QY 1670 ATGAAAAATTCATGGAAGCAATGGG 1694
|||||
Db 61 ATGAAAAATTCATGGAAGCAATGGG 85
|||||
RESULT 5
BU265304
LOCUS   603508640F1 CSEQCHN52 Gallus gallus CDNA clone CHEST437010 5', mRNA
DEFINITION
ACCESSION
BU265304
VERSION 603508640F1 GI:25536254
KEYWORDS
SOURCE   Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 759)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE   A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..759
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST437010"
/dev_stage="22"
/lab_host="DH10B"
/clone_lib="CSEQCHN52"
/notes="Organ: limbs; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
ORIGIN
Query Match 3.5%; Score 84; DB 5; Length 759;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1611 GCCTACAGACAGAAAGATGCGATTTAACGCTACTTCGAAATAGAGAAAAATGAGAACTA 1670
|||||
Db 1 GCCTACAGACAGAAAGATGCGATTTAACGCTACTTCGAAATAGAGAAAAATGAGAACTA 60
|||||
QY 1671 TGAATAATTCATGGAAGCAATGGG 1694
|||||
Db 61 TGAATAATTCATGGAAGCAATGGG 84
|||||
RESULT 6
BU296697
LOCUS   603741965F1 CSEQCHN56 Gallus gallus CDNA clone CHEST640b17 5', mRNA
DEFINITION
ACCESSION
BU296697
VERSION 603741965F1 GI:25746333
KEYWORDS

```

REFERENCE	1	(bases 1 to 885)
AUTHORS	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
TITLE	A Comprehensive Collection of Chicken cDNAs	
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	
MEDLINE	22335534	
PUBMED	12445392	
COMMENT	Contact: Simon Hubbard Department of Biomolecular Sciences University of Manchester Institute of Science and Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK Tel: 01612008930 Fax: 01612360409 Email: Simon.Hubbard@umist.ac.uk.	
FEATURES	Location/Qualifiers	
source	1..885	
	/organism="Gallus gallus"	
	/mol_type="mRNA"	
	/strain="Compton Line 151"	
	/db_xref="taxon:9031"	
	/clone="CHEST14c20"	
	/sex="Female"	
	/dev_stage="adult"	
	/lab_host="DH10B"	
	/clone_lib="CSEQCHL18"	
	/note="Organ: small intestine; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; Modification of pBluescript II KS(+) [Stratagene] vector to accommodate cDNA produced with the T-trimmed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BglII and BamHI sites [5'ggccgcgtcgacgccgcttcgcaaaaagaag][5'aattcttttcttgatccgggtcgacgc]"	
ORIGIN		
Query Match	3.3%; Score 79.4; DB 5; Length 885;	
Best Local Similarity	98.8%; Pred.No.9.2e-08;	
Matches	80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1614	TACAGACAGAAGATGGCATTTAACGGTACTTGGAATAAGAGAAAAATGAGAACTATGA 1673
Dd	1	TTACAGACAGAAGATGGCATTTAACGGTACTTGGAATAAGAGAAAAATGAGAACTATGA 60
QY	1674	AAAATTTCATGGAAGCAATGGG 1694
Dd	61	AAAATTTCATGGAAGCAATGGG 81
RESULT 8		
BU234099		
LOCUS	BU234099	603792609F1 CSEQCHN24 Gallus gallus cdna clone CHES758b4 5', mRNA sequence.
DEFINITION		
ACCESSION	BU234099.1	GI:25478348
VERSION		
KEYWORDS		EST.
SOURCE		Gallus gallus (chicken)
ORGANISM		Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; 1 (bases 1 to 814)		
REFERENCE		
AUTHORS	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
TITLE	A Comprehensive Collection of Chicken cDNAs	
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	
MEDLINE	22335534	
PUBMED	12445392	
COMMENT	Contact: Simon Hubbard	

SOURCE	Gallus gallus (chicken)	
ORGANISM	Gallus gallus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; 1 (bases 1 to 829)		
REFERENCE		
AUTHORS	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
TITLE	A Comprehensive Collection of Chicken cDNAs	
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	
MEDLINE	22335534	
PUBMED	12445392	
COMMENT	Contact: Simon Hubbard Department of Biomolecular Sciences University of Manchester Institute of Science and Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK Tel: 01612008930 Fax: 01612360409 Email: Simon.Hubbard@umist.ac.uk.	
FEATURES	Location/Qualifiers	
source	1..829	
	/organism="Gallus gallus"	
	/mol_type="mRNA"	
	/strain="Compton Line 151"	
	/db_xref="taxon:9031"	
	/clone="CHEST640b17"	
	/sex="Female"	
	/dev_stage="adult"	
	/lab_host="DH10B"	
	/clone_lib="CSEQCHN56"	
	/note="Organ: small intestine; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."	
ORIGIN		
Query Match	3.4%; Score 81; DB 5; Length 829;	
Best Local Similarity	100.0%; Pred.No.3.7e-08;	
Matches	81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1614	TACAGACAGAAGATGGCATTTAACGGTACTTGGAATAAGAGAAAAATGAGAACTATGA 1673
Dd	11	TACAGACAGAAGATGGCATTTAACGGTACTTGGAATAAGAGAAAAATGAGAACTATGA 70
QY	1674	AAAATTTCATGGAAGCAATGGG 1694
Dd	71	AAAATTTCATGGAAGCAATGGG 91
RESULT 7		
BU123280		
LOCUS	BU123280	603003274F1 CSEQCHL18 Gallus gallus cdna clone CHEST14c20 5', mRNA sequence.
DEFINITION		
ACCESSION	BU123280.1	GI:25333903
VERSION		
KEYWORDS		EST.
SOURCE		Gallus gallus (chicken)
ORGANISM		Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; 1 (bases 1 to 814)		
REFERENCE		
AUTHORS	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
TITLE	A Comprehensive Collection of Chicken cDNAs	
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	
MEDLINE	22335534	
PUBMED	12445392	
COMMENT	Contact: Simon Hubbard	

Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

1. 814  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="ChEST758b4"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN24"  
/note="Organ: heads; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

FEATURES  
source

Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
source

1. 1080  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="ChEST153f19"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHL18"  
/note="Organ: small intestine; Vector: pBluescript II  
KS(+); Site 1: EcoRI; Site 2: NotI; Modification of  
pBluescript-II KS(+) [Stratagene] vector to accommodate  
cDNA produced with the T-trimmed protocol (construction of  
uni-directionally cloned cDNA libraries from messenger RNA  
for improved 3' end DNA sequencing by Glenn Fu, et al.  
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with  
NotI and EcoRI. Ligate in double stranded adaptor  
containing BsgI and BamHI sites  
[5'ggcgcgcgagcccgatccggatccggggctgcacgc]  
[5'attcttttttggatccggggctgcacgc]"

ORIGIN

Query Match 3.3%; Score 78.8; DB 5; Length 1080;  
Best Local Similarity 97.6%; Pred. No. 1.3e-07;  
Matches 80; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1613 CTACAGACAGAAAGATGCGATTAAACGGTACTTCGAAATAAGAGAAAATGAGAACTATG 1672

Db 7 CGAGAGACAGAAAGATGCGATTAAACGGTACTTCGAAATAAGAGAAAATGAGAACTATG 66

Qy 1673 AAAAATTCATGGAAGCAATGGG 1694

Db 67 AAAAATTCATGGAAGCAATGGG 88

Qy 1616 CAGACAGAAAGATGCGATTAAACGGTACTTCGAAATAAGAGAAAATGAGAACTATGAAA 1675

Db 1 CAGACAGAAAGATGCGATTAAACGGTACTTCGAAATAAGAGAAAATGAGAACTATGAAA 60

Qy 1676 AATTCATGGAAGCAATGGG 1694

Db 61 AATTCATGGAAGCAATGGG 79

RESULT 9  
BUI23336  
LOCUS 603149475F1 CSEQCHL18 1080 bp mRNA linear EST 25-NOV-2002  
DEFINITION 603149475F1 CSEQCHL18 Gallus gallus cDNA clone ChEST153f19 5', mRNA  
sequence.  
ACCESSION BUI23336  
VERSION BUI23336.1 GI:253333972  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 1080)  
AUTHORS Boardman,P.E., Sang-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 22335534  
PUBMED 12445392

COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409

FEATURES  
source

1. 426  
/organism="Gallus gallus"

CD739695 426 bp mRNA linear EST 26-JUN-2003  
LOCUS 4028031 1GAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA  
DEFINITION clone 1GAL\_20B02 5', mRNA sequence.  
ACCESSION CD739695  
VERSION CD739695.1 GI:32290544  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 426)  
AUTHORS Min,W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van  
Tassel,C. and Han,J.Y.  
TITLE Chicken intestinal lymphocyte EST database as a resource for the  
analysis of mucosal immune function  
JOURNAL Unpublished (2003)  
COMMENT Contact: Hyun S. Lillehoj  
Animal Parasite Diseases Laboratory  
Animal and Natural Resources Institute, USDA  
Bldg.1043, BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048771  
Fax: 3015045103

Email: lilleho@anri.barc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim alt -, -trim fasta. Vector identified  
by cross match using options -minmatch 12 -minscore 12  
Plate: 20 row: B column: 02  
Seq primer: ATTTAGTGACACTATAG  
High quality sequence stop: 426.  
Location/Qualifiers  
1. 426  
/organism="Gallus gallus"







Query Match		2.6%;	Score 61.4;	DB 5;	Length 885;
Best Local Similarity		96.1%;	Pred. No. 0.0018;		
Matches		74;	Conservative	0;	Mismatches 1; Indels 2; Gaps 1;
Qy	1618	GACAGAAAGATGGCATTTAACGGTACTTGGAAATAGAGAAAAATGAGAACTATGAAAAA	1677		
	8	GAGAGAAAGATGGCATTTAACGGTACTTGGAAAT--AGAAAAATGAGAACTATGAAAAA	65		
Db	1678	TTTCATGGAAGCAATGGG	1694		
	66	TTTCATGGAAGCAATGGG	82		
RESULT 15					
LOCUS					
DEFINITION					
BX276255 AGENAE Gallus gallus multi-tissues normalized library					
(gcag) Gallus gallus cDNA clone gcag0008c.c.24 5prim, mRNA					
sequence.					
ACCESSION					
BX276255					
VERSION					
BX276255.1 GI:28598746					
KEYWORDS					
EST.					
SOURCE					
Gallus gallus (chicken)					
ORGANISM					
Gallus gallus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;					
Phasianinae; Gallus.					
REFERENCE					
1 (bases 1 to 443)					
AUTHORS					
Herault,F., Le Meuth-Metzinger,V., Desert,C., Retout,E., Piumi,F.,					
Klopp,C. and Douaire,M.					
TITLE					
Construction and primary characterization of chicken normalized					
multi-tissue cDNA libraries					
JOURNAL					
Unpublished (2003)					
COMMENT					
Contact: Douaire M					
INRA, UMR INRA-ENSAR Genetique Animale					
65, rue de Saint-Brieuc, RENNES cedex, 35042, FRANCE					
Tel: +33 (0) 2.23.48.54.63					
Fax: +33 (0) 2.23.48.54.70					
Email: Madeleine.Douaire@roazhon.inra.fr					
Sequence cleaned of vector, adaptor and repetitions. Contact us					
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this					
sequence.					
Plate: 0008 row: c column: 24					
Seq primer: M13R.					
FEATURES					
source					
1..443					
Location/Qualifiers					
/organism="Gallus gallus"					
/mol_type="mRNA"					
/db_xref="taxon:9031"					
/clone="gcag0008c.c.24"					
/tissue_type="multi-tissues"					
/dev_stage="from embryos to adults"					
/lab_host="DH10B"					
/clone_lib="AGENAE"					
library (gcag)"					
/note="Vector: pT73D-pac; tissues: brain, embryos,					
kidney, multi-tissues, muscle, pancreas, skin, testis,					
liver, adipose tissue, granulosa, utero-vaginal gland,					
oviduct, small follicle, ovary, hypothalamus, pituitary					
gland, ileon, jejunum, caecum, duodenum, spleen,					
fabricius gland, bone marrow, thymus, hematopoietic					
progenitor cells. Clone distribution : AGENAE Resource					
centre. Francois PIUMI, Francois.Piumi.inra.fr, INRA, CEA					
Radiobiologie et Etude du genome (LREG), Domaine de					
Vilvert, 78352, Jouy-en-Josas cedex, FRANCE"					
ORIGIN					
Query Match		2.5%;	Score 60.4;	DB 5;	Length 443;
Best Local Similarity		98.4%;	Pred. No. 0.0026;		
Matches		61;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	2320	TAGCGGTGAATGTGATGAAAAAGTTAGGAGCCCGACGATATCTGAAGCTCATTTC	2379		

Db	31	TGGGCGTGAATGTGATGAAAAAGTTAGGAGCCCGACGATATCTGAAGCTCATTTC	90
Qy	2380	AG 2381	
Db	91	AG 92	
Search completed: January 14, 2005, 12:21:35			
Job time : 8253.18 secs			

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 07:35:21 ; Search time 1488.19 Seconds  
(without alignments)  
10676.925 Million cell updates/sec

Title: US-10-099-663-2

Perfect score: 336

Sequence: 1 attattatttttcattgata.....agacagaagatggcattta 336

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					
C 1	42.8	12.7	88688	9	AC087429 Homo sapi
C 2	42.8	12.7	187003	3	AC023353 Homo sapi
C 3	41.8	12.4	252420	3	AB014841 Plasmodiu
C 4	41.4	12.3	1663	3	AK116155
C 5	41.4	12.3	99855	8	OSJN001175
C 6	41.4	12.3	121130	8	OSJN00292
C 7	40.6	12.1	211947	2	AL928605
C 8	40.4	12.0	60756	2	AL583823
C 9	40.4	12.0	131708	9	AL354710
C 10	40.4	12.0	174294	2	AC026155
C 11	40.2	12.0	110000	10	AB014175
C 12	40.2	12.0	201783	10	AC114410
C 13	39.8	11.8	177655	2	CR318618
C 14	39.8	11.8	217375	5	BX537337
C 15	39.6	11.8	162605	2	AC068698
C 16	39.6	11.8	164275	9	AC005157
C 17	39.6	11.8	230585	10	AC110259
C 18	39.4	11.7	393	8	YSCMTORIE
C 19	39.4	11.7	396	8	MIECORIS

C	20	39.4	11.7	130574	9	AL160174
C	21	39.4	11.7	141668	10	BX001028
C	22	39.4	11.7	172945	9	AC067719
C	23	39.4	11.7	193306	2	AC025512
C	24	39.4	11.7	194638	2	AC112913
C	25	39.4	11.7	238301	2	AC094376
C	26	39	11.6	2933	6	CQ597407
C	27	39	11.6	9124	2	AC020330
C	28	39	11.6	86258	5	BX842699
C	29	39	11.6	108924	3	AC005889
C	30	39	11.6	177997	3	AC092492
C	31	39	11.6	270775	3	AE003624
C	32	38.8	11.5	1031	3	TETHIS2A1A
C	33	38.8	11.5	109465	8	AP006106
C	34	38.8	11.5	139401	2	AC124406
C	35	38.8	11.5	210359	9	HS4330K20
C	36	38.8	11.5	212445	10	AC108419
C	37	38.6	11.5	165873	5	AL954323
C	38	38.6	11.5	207372	10	AL731851
C	39	38.6	11.5	214950	2	AC110406
C	40	38.6	11.5	246528	2	AC128573
C	41	38.4	11.4	66569	2	AC103837
C	42	38.4	11.4	154433	10	AL732447
C	43	38.4	11.4	166860	10	AC115924
C	44	38.4	11.4	177967	10	AC122816
C	45	38.4	11.4	189019	2	AC129426

#### ALIGNMENTS

RESULT 1	AC087429/c	AC087429	88688 bp	DNA	linear	PRI 11-APR-2002
LOCUS	AC087429	Homo sapiens chromosome 3 clone 2063K18 map 3p,	complete sequence.			
DEFINITION	AC087429	HTG.				
ACCESSION	AC087429.2	GI:20136892				
VERSION	HTG.					
KEYWORDS	HTG.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 88688)					
AUTHORS	Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.					
TITLE	Chromosome 3p genomic sequence					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 88688)					
AUTHORS	Liu, B., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.					
TITLE	Submitted (05-JAN-2001)					
JOURNAL	Direct Submission					
REFERENCE	2. Institute of Genetics, Chinese Human Genome Center at Shanghai					
AUTHORS	3 (bases 1 to 88688)					
REFERENCE	Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Hu, S., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.					

Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
Yu, J. and Yang, H.  
Direct Submission  
Submitted (11-APR-2002) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P. R. China  
On Apr 11, 2002 this sequence version replaced gi:12039240.  
-----Genome Center  
Center: Beijing Center  
Center code: Beijing  
Website: http://hgsc.igtp.ac.cn  
http://www.genomics.org.cn  
Contact: hgc@igtp.ac.cn  
----- Project Information  
Center project name: 11x project  
Center clone name: 2063K18  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator; ET 55% of reads  
Chemistry: Dye-terminator Big Dye; 45% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 89950 bases at least Q40  
Consensus quality: 90370 bases at least Q30  
Consensus quality: 90443 bases at least Q20  
Insert size: 88688; sum-of-contigs  
Quality coverage: 7.47x in Q20 bases; sum-of-contigs

On Apr 11, 2002 this sequence version replaced gi:12039240.

JOURNAL  
COMMENT

Submitted (11-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jul 15, 2000 this sequence version replaced gi:7235346.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information  
Center project name: H.NH0775C23  
----- Summary Statistics  
Sequencing vector: M13; 57%  
Sequencing vector: plasmid; 43%  
Chemistry: Dye-terminator; ET 57% of reads  
Chemistry: Dye-terminator Big Dye; 43% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 172430 bases at least Q40  
Consensus quality: 176603 bases at least Q30  
Consensus quality: 178873 bases at least Q20  
Insert size: 189000; agarose-fp  
Insert size: 184303; sum-of-contigs  
Quality coverage: 3.50 in Q20 bases; agarose-fp  
Quality coverage: 3.66 in Q20 bases; sum-of-contigs

----- Project Information  
Center project name: H.NH0775C23  
----- Summary Statistics  
Sequencing vector: M13; 57%  
Sequencing vector: plasmid; 43%  
Chemistry: Dye-terminator; ET 57% of reads  
Chemistry: Dye-terminator Big Dye; 43% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 172430 bases at least Q40  
Consensus quality: 176603 bases at least Q30  
Consensus quality: 178873 bases at least Q20  
Insert size: 189000; agarose-fp  
Insert size: 184303; sum-of-contigs  
Quality coverage: 3.50 in Q20 bases; agarose-fp  
Quality coverage: 3.66 in Q20 bases; sum-of-contigs

----- Project Information  
Center project name: H.NH0775C23  
----- Summary Statistics  
Sequencing vector: M13; 57%  
Sequencing vector: plasmid; 43%  
Chemistry: Dye-terminator; ET 57% of reads  
Chemistry: Dye-terminator Big Dye; 43% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 172430 bases at least Q40  
Consensus quality: 176603 bases at least Q30  
Consensus quality: 178873 bases at least Q20  
Insert size: 189000; agarose-fp  
Insert size: 184303; sum-of-contigs  
Quality coverage: 3.50 in Q20 bases; agarose-fp  
Quality coverage: 3.66 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source

Location/Qualifiers  
1. 88688  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3p"  
/clone="2063K18"

ORIGIN

Query Match 12.7% Score 42.8; DB 9; Length 88688;  
Best Local Similarity 53.6%; Pred. No. 1.2; Indels 0; Gaps 0;  
Matches 89; Conservative 0; Mismatches 77;  
Qy 1 ATTATTATTTTCATTAGATACCGGTTTTTACTACAACTCAAATAAGATGAACAAGATG 60  
Db 1637 ATTAACATTAAACAAATTAGTAGCCTTTATATACCTTTTAACACGACATTTAGAGACT 1578  
Qy 61 AATGGGTAGTACTGTTTATAAGACAGATATAAGATACATCATCATTTTGAGGCAA 120  
Db 1577 TAAGAGTAAGTCCATGTATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1518  
Qy 121 TAAGGGAGGGAGAGATTTCAGCAACACAGTGTGCTTACAAAGTGGAAAA 166  
Db 1517 TAAGACACTAAATACCTTAGAATAAATTTAATAAATAAGTACAAA 1472

RESULT 2

AC023353/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 20 clone RP11-775C23, WORKING DRAFT  
SEQUENCE, 28 unordered pieces.  
AC023353  
VERSION  
AC023353.3 GI:9211518  
KEYWORDS  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Waterston, R.H.  
TITLE  
The sequence of Homo sapiens clone  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 187003)  
Waterston, R.H.  
Direct Submission

1 1408: contig of 1408 bp in length  
1409 1508: gap of unknown length  
1509 3342: contig of 1834 bp in length  
3343 3442: gap of unknown length  
3443 5329: contig of 1887 bp in length  
5329 5429: gap of unknown length  
5429 8611: contig of 3182 bp in length  
8611 8711: gap of unknown length  
8711 11044: contig of 2333 bp in length  
11044 11144: gap of unknown length  
11144 15105: contig of 3961 bp in length  
15105 15206: gap of unknown length  
15206 17732: contig of 2527 bp in length  
17732 17832: gap of unknown length  
17832 21791: contig of 3959 bp in length  
21791 21991: gap of unknown length  
21991 23547: contig of 1656 bp in length  
23547 23648: gap of unknown length  
23648 28121: contig of 4474 bp in length  
28121 28221: gap of unknown length  
28221 34018: contig of 5797 bp in length  
34018 34118: gap of unknown length  
34118 38469: contig of 4351 bp in length  
38469 38569: gap of unknown length  
38569 44271: contig of 5702 bp in length  
44271 44371: gap of unknown length  
44371 51457: contig of 7086 bp in length  
51457 51557: gap of unknown length  
51557 58979: contig of 7422 bp in length  
58979 59079: gap of unknown length  
59079 65659: contig of 6580 bp in length  
65659 65759: gap of unknown length  
65759 74077: contig of 8318 bp in length  
74077 74177: gap of unknown length  
74177 82579: contig of 8402 bp in length  
82579 82679: gap of unknown length  
82679 92223: contig of 9544 bp in length  
92223 92323: gap of unknown length  
92323 99655: contig of 7332 bp in length  
99655 99755: gap of unknown length







genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (<http://genemark.biology.gatech.edu/GeneMark/>), tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the complete sequence against NCBI none redundant protein database (nr) (<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.

FEATURES

source

1. .99855  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/sub\_division="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:39947"  
/chromosome="4"  
/clone="OSJNBa0064D20"  
/clone\_1db="CGI-OSJNBa"  
complement (3866. .6899)  
/gene="OSJNBa0064D20.1"  
complement (join(3866. .5543,5638. .6485,6573. .6999))  
/gene="OSJNBa0064D20.1"  
/codon\_start=1  
/protein\_id="CAE04217.1"  
/db\_xref="GI:32489561"  
/db\_xref="TrEMBL:Q7X84"

gene

CDS

/translation="MARGSAPLDGSLPPSRIVSROAGLPFRFMPESAIGREIVSLG  
EGCAPRYGSRVFEIPEFGMAGLVPPESPFDVLEFDLQMAHLTPNAVMTLIFTH  
LCMEFTGPRGPVLRNYPVCLVRKNDWDKSDWFTPLDEARLRLRSQPPAQASWVA  
PVLGDGWDAPEDFKIIVQRVLNLSVEASLI PQGLPLCDSPDRASITITIMQAVAS  
EERAPGAGGSGRGEOSTPGGRASGPRDGGGSGRPAAGRKRGKEGTPPPSP  
KGGGAARASRRSEGAFTPTSQPERGKKRLRKMGKTEPSRGNLI SPPWFSNRPSP  
EIPSPRSKSGSAEADPATTEARRSDREAAADRLREAEAREAREARATEEA  
AREAGLRQAEATTSSEATRETDQEAADRLREAEAAQAVRARQAEAAAREEARL  
QABATTISEATREDAAGSLGPTCDQKPGFDIPESGTSISGPSRASSRRRLF  
PTFPAPLSAPHLALQALAVANTVLDGLSAQVEALQAEALDAAWAVEGRSVEA  
MVVGRHRRHSELEARRKVLGETAKEVEEERGAALITATTMNEAQTLLRQYGSW  
EALGLDAARGLVDAARAAAREAAETEVAARREREAALREARAEAAACRAVEDLSLR  
DREAVALREATTAEACAEESALRLREDALTRERALEAEAAACRAVEDLSLR  
EAAREQACNLREGARAAALNORAAELREARELDARSGGAAGESDLTARLAA  
AEHTIADLQGLSSAGEVEALRAGEVGPFRDVA SRLDRAGRQGLWGRRTAKIA  
ANGGGLAQLSEIAGTILQRLPEELRTIKSSSRDLA"  
complement (10918. .12816)  
/gene="OSJNBa0064D20.2"  
complement (join(10918. .11100,11189. .11344,11862. .11979,  
12107. .12816))  
/gene="OSJNBa0064D20.2"  
/codon\_start=1  
/protein\_id="CAE04218.2"  
/db\_xref="GI:38346571"  
/db\_xref="TrEMBL:Q7X902"

gene

CDS

/translation="MGRFRCCLIDDLQGLLILHWPGRGAVOTPSPPATRLSL  
PDALLTILSLKGRQVQTSVLKRWRLHWSVPCLDIDQREFAAASENWALSRL  
EKFEFADNVLYATRCSPAKIDTFRURI CDRYHSRSDTDWRIRGLKCSPREFLH  
FDYRDSYLLEMHKSGSYGCLTKLNTVSLHCEPMEHTTVCITLLELNRCSLY  
LQBITPKLNLVLHGPVMPSLVRASIRFVVPCKNCFLNVTLSLELSGFRM  
VRLFEQVLFHAPKNTLLDRCRPSYNNELLRHLLONS PNLEKLTVCCKCFKSGSL  
ENRKSQHNQVNCRLKSTIYIKDIDVDRELVDLLDVSGHLPKNTLTKI"  
complement (23800. .27420)  
/gene="OSJNBa0064D20.3"  
complement (join(23800. .24075,24708. .24800,24919. .24998,  
25802. .25967,26575. .27420))  
/gene="OSJNBa0064D20.3"  
/codon\_start=1  
/protein\_id="CAE04219.2"  
/db\_xref="GI:38346572"  
/db\_xref="TrEMBL:Q7X801"

gene

CDS

/translation="MEFVAKRATGGGSAADVRLSSILPDALLHAVMSSLPARQMVQ  
TCVLSRWVHLWSVNLNDSREFLPIYDRWQMEDFTTNLMPHFHAPTLDAFSIR  
ADVAGKHGRHVDWRIRCGIKYCPRLVIDAVTVDSRVLDPDLSGSCFLGRHLISYV  
ALHSGFARQVRQSCVPLRCLERCLTKESHIESSTLNLRLVIEDSGSGSDSLAIPR  
LASRLVAFELPHYKNGVSLNGANSILVEASVAVKGRSTSPGEAMLLGLPSVLSLEL  
KGLQELVFLVGLVIQKSWRLAFTTLVEISEFOAILHEKFDKSGSFHNRLTSLDNCF  
QAGDADVAALDSWQVPRPILSLQRLEDKVTVFQEPDELSMFYNLKRACCSIASA  
NVYGAWESEHNDHGDRRPSLAAAPSQSPAAGLRAAPPIVSTRGRGRLHGEARQR

gene

CDS

CCRRSLFAGLSSPHVPAGRRRHIGGDIVGEDAVAQ"  
complement (29612. .32646)  
/gene="OSJNBa0064D20.4"  
complement (join(29612. .29827,29918. .30088,30550. .30590,  
31036. .31096,31762. .32646))  
/gene="OSJNBa0064D20.4"  
/codon\_start=1  
/protein\_id="CAE04220.2"  
/db\_xref="GI:38346573"  
/db\_xref="TrEMBL:Q7X877"  
/translation="MDRLSALPDGLLHAWMFLPAROMAQTCLVSKRWHLWRSVPSL  
NLDIRFERNLSKEEDDDDETGMKMDFTANILMFHAPTLDTFTCTGACVQAAVGN  
RSRRNRNDVDYPAVANRSRDVGRWIRGGIKYCPRLVDHVAVPSGSCRPPLDLS  
GSCFRRLRLRSLFALDSGFARQVRDSCVLRCLDLHRLHIEFHIESSTLUNKLVIE  
GCIGCSLAI SAPRLASLCLDLSYGAYKNGVSLNTI SLVEASVTLNVFQISPEGEAM  
LLCGLFNTVNLLEAIAHAKLGLVLA KSAVGSKRGGRERKPKYCCCNVNEKVI LHE  
KDFKPSFNRLRTLSDDCFOGMDLKEFKALGRLLKPCPNLEKLTQLQHCWFLSGST  
QRAEMETRTSGLCQSQQVTFHCQKLKLEIKHDSLCCDHLQLFQLMGFWFKDLKKA  
TILITKI"  
complement (37808. .41919)  
/gene="OSJNBa0064D20.5"  
complement (join(37808. .37912,38701. .38871,38985. .39135,  
41339. .41919))  
/gene="OSJNBa0064D20.5"  
/codon\_start=1  
/protein\_id="CAE04221.2"  
/db\_xref="GI:38346574"  
/db\_xref="TrEMBL:Q7X6F3"

gene

CDS

/translation="MGRPRRAPATVLEINVTPSPYPPGHVGYTMPDLIITSRARW  
LTRHLSHAWLDGGFGEOLGDGCPLELRLALRRCCAMAPFRIRCGSLRTLVLHTGC  
GGDAGGGDEBEETLIVISAPRLASVRKITSYACRHGVSPDGSSTABLSVASIRVG  
RRRRALUPTGVEAVLLAGMNVITLSSKVRQDRGTVSLGVNSLVSASIDSNQCM  
SPSGAMLLGALFSAITLRLKRAMALIDELDKPFLPNLNLRLSLHCCLRDKNLS  
DRFKALGRLLQKSLNLEKLTQLQDFWASWCRNTSSVLGKIQLVLLFTLPIILHAVSDL  
N"  
complement (45448. .46707)  
/gene="OSJNBa0064D20.6"  
complement (join(45448. .45516,45762. .46130,46270. .46707))  
/gene="OSJNBa0064D20.6"  
/codon\_start=1  
/protein\_id="CAE04222.2"  
/db\_xref="GI:38346575"  
/db\_xref="TrEMBL:Q7X5X6"

gene

CDS

/translation="MADRTTATMI PAASSSRKRARVTTGAGDGGGGGEGRLGELPD  
ELLUSILCLTTRQAVQTSVLRSRWRHLWRSTPFDVDAEFAFPSPSAPMLHGRG  
STDPMERLRGFTARLLMFAAPVLDARLVA TPFRHRADEVSAFAGDLSGLCPAL  
VDMELDRCKCFHELSSATLRLSAMESCLMWRPSGTNGDRTVSVVAPRLAYRLITF  
RGDCCKVFRFSGDSISVSIIRGGPNLNLFRLLRMPNVITLRLSGFGPTSVSARME  
RKFPFPPOSDNLALG"  
47424. .50668  
/gene="OSJNBa0064D20.7"  
join(47424. .47697,48191. .48465,48691. .48855,48923. .49012,  
50294. .50668)  
/gene="OSJNBa0064D20.7"  
/codon\_start=1  
/protein\_id="CAE04223.2"  
/db\_xref="GI:38346576"  
/db\_xref="TrEMBL:Q7X8C0"

gene

CDS

/translation="MGGGGGGRNVNDHDDDDDDWELAAAGALADTVTLVLVKGVGSGKS  
ATANSILGDEAFESKSYAGVTQCKKSTTVQDGLITINVIDTGPLFDMDIKAE  
VRREIVKCDMAKDCIGIHAMLVFSA TSRSFCEDEKTIETLKSFGDKILDHMLVFT  
GDEVGETISWKNMLSDSAPTLYLDILKLFENRVL FENKTSSTQDRQQRKMKMLDAVD  
FVVSNGHKPFPSNQLFTQIQEVHHRQKDANSEVSSMOETDSYISLITKMOVEEKUNG  
ILRMFQQLLKEQKQEDIONEMTKAILRSEEDIRRLRLSLKAEQSNNAEENKRF  
ESEKASKEQEKQTEAEIQKLKEKMEKREEREERIRLRDLDEKEREREKQSGCIL  
"  
complement (51830. .56465)  
/gene="OSJNBa0064D20.8"  
complement (join(51830. .52739,52843. .52970,54357. .54818,  
56012. .56229,56348. .56465))  
/gene="OSJNBa0064D20.8"  
/codon\_start=1  
/protein\_id="CAE04224.2"

gene

CDS

complement (51830. .56465)  
/gene="OSJNBa0064D20.8"  
complement (join(51830. .52739,52843. .52970,54357. .54818,  
56012. .56229,56348. .56465))  
/gene="OSJNBa0064D20.8"  
/codon\_start=1  
/protein\_id="CAE04224.2"







\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 211947: contig of 211947 bp in length.

FEATURES  
source  
1..211947  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-139P14"  
/clone\_lib="RPCI-23"  
misc\_feature  
1..211947  
/note="assembly\_fragment:03130  
clone\_end:SP6  
vector\_side:left"

ORIGIN  
Query Match 12.1%; Score 40.6; DB 2; Length 211947;  
Best Local Similarity 53.5%; Pred. No. 4.2;  
Matches 85; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
Qy 22 CCGGTTTTTACTCAACTCAATAAGATGAACAGAAATGAATGGTTAGTGACTGTTTAT 81  
Db 194980 CTGATTATTTTCATCAAGGAAGAAAGAAACAAAGGAGTAAGGAGAAGTCAGAAAGAA 195039  
Qy 82 AAAGAAGAGTAAATAAGATCTATCATCTTTTGAGGCAATAAGCGGGAGAGATTCCAGC 141  
Db 195040 AAAGAAGATGAAGAAAGAAAGAAAGAAAGAGTAAGGAGGAGCGGACGGAGGAGG 195099  
Qy 142 AACAGCTGCTTACAGTGAACAAACAGTTAAACTAAA 180  
Db 195100 TAGGAGTGGGGGAAAGAAAGAGTTAAAAAAGAAAGAA 195138

RESULT 8  
AL583823\_3/c  
WPCOMMENT  
Sequence split into 4 fragments LOCUS AL583823 Accession AL583823  
Fragment Name Begin End  
AL583823\_0 1 110000  
AL583823\_1 100001 210000  
AL583823\_2 200001 310000  
AL583823\_3 300001 360756  
Continuation (4 of 4) of AL583823 from base 300001 (AL583823 Homo sapiens chromosome 9 c  
Query Match 12.0%; Score 40.4; DB 2; Length 60756;  
Best Local Similarity 55.8%; Pred. No. 5.2;  
Matches 77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 197 ACAAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCCT 256  
Db 33572 AAAAGATGATGCCCTCAACTAGGTTTGACACAGTGATTTTAGCATGATGAATTAACCT 33513  
Qy 257 GATTAAGCTGTTTCAATAATCTCTTTGCAAGCTCTGCTACTTACCAGAAGTCGCTAC 316  
Db 33512 GGTTAGCTTTTAAATAATTCATTTCCCAAGCTCTTACTTCTTAAAAAAGAGTCAGTTT 33453  
Qy 317 AGACAGAAAGATGGCAAT 334  
Db 33452 AGAAGAAAGAGATATT 33435

RESULT 9  
AL354710  
LOCUS AL354710 131708 bp DNA linear PRI 30-JAN-2002  
DEFINITION Human DNA sequence from clone RP11-65N13 on chromosome 9, complete  
sequence.  
ACCESSION AL354710  
VERSION AL354710.17 GI:18476556

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 131708)  
Lloyd, D.  
Direct Submission  
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:18375790.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-65N13 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone  
RP11-65N13 It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true right end of clone RP11-65N13 is at 131708 in this  
sequence. The true left end of clone RP11-184B22 is at 91097 in  
this sequence. The true right end of clone RP11-366020 is at 2000  
in this sequence.  
Location/Qualifiers  
source  
1..131708  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-65N13"  
/clone\_lib="RPCI-11.1"  
68692..68732  
/notes="Sequence from overlapping clone RP11-180F6  
(AL583823). Assembly confirmed by restriction digest."  
69053..69060  
/notes="Sequence from overlapping clone RP11-180F6  
(AL583823) and RP11-233N7 (AL445234). Assembly confirmed  
by restriction digest."  
80432..81341  
/notes="Sequence from overlapping clone AC026155 sequenced  
by BCM and RP11-180F6 (AL583823). Assembly confirmed by  
restriction digest."  
89032..89102  
/notes="Sequence from overlapping clone RP11-180F6  
(AL583823). Assembly confirmed by restriction digest."

FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

ORIGIN

Query Match 12.0%; Score 40.4; DB 9; Length 131708;  
Best Local Similarity 55.8%; Pred. No. 4.9;  
Matches 77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;



Db 117389 GGTAGCTTTTAAAAATTCATTCCCAAGTCTTACTTCTTAAAAAAGAGTCAGTGTTTT 117330  
Qy 317 AGACAGAAAGATGCCATT 334  
Db 117329 AGAAGAAAAAGATATT 117312

## RESULT 11

AE014175 2/c

## WPCOMMENT

Sequence split into 4 fragments LOCUS AE014175 Accession AE014175

Fragment Name Begin End

AE014175\_0 1 110000

AE014175\_1 100001 210000

AE014175\_2 200001 310000

AE014175\_3 300001 404829

Continuation (3 of 4) of AE014175 from base 200001 (AE014175 Mus musculus piebald deleti

## Query Match

Best Local Similarity 48.9%; Score 40.2; DB 10; Length 110000;

Matches 108; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 26 TTTTCTACTACAACTCAAAATAGATGAACAGATGATGGTTAGTACTGTTTATATAAG 85

Db 15331 TTTTATAGCTAAAAATTTAGAAAAATGCTTCATTCTCTGTTATTTGCTATTATAAATTAT 15272

Qy 86 AAGATGATTAAGATCTATCATCATTTGAGCAATRAAGGAGGAGGAGATTCAGCAAC 145

Db 15271 CAGACATGTTTCTTTTCAGTCTAGAGTGAGTTATTTGAGAGGAGAGACACACAGTG 15212

Qy 146 AGTGTGCTTACAGTGGAAAAACAAGTTAAACTAAAGTGACCCCTCTCTGACAAATCA 205

Db 15211 AGTGTGAGAACAGTGCAGCACCATGACCCCTGGCAAAAGCTTTCATTATAAGAACA 15152

Qy 206 ATGCCACAGTGACGTTTAGCCACCCACATCATCATGTA 246

Db 15151 GGGAGAGGGAGGAGTGAAGGGCGTGCCACATCATTCACA 15111

## RESULT 12

AC114410/c

## LOCUS

DEFINITION Mus musculus, clone RP23-151K8, complete sequence. ROD 09-SEP-2003

ACCESSION AC114410

VERSION AC114410.9 GI:34536773

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 2 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 3 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 4 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 5 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 6 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

REFERENCE 7 (bases 1 to 201783)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

JOURNAL Unpublished

TITLE Mus musculus, clone RP23-151K8

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

## Direct Submission

Submitted (08-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 201783)

Direct Submission

Birren, B., Nusbaum, C. and Lander, E.

## Direct Submission

Submitted (15-AUG-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 201783)

Direct Submission

Birren, B., Nusbaum, C. and Lander, E.

## Direct Submission

Submitted (09-SEP-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 9, 2003 this sequence version replaced gi:33667216.

All repeats were identified using RepeatMasker:

Snit, A.P.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L23439

Center clone name: 151\_K\_8

## FEATURES

source

1..201783

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP23-151K8"

/clone\_lib="RPCI-23 Female Mouse BAC"

misc\_feature

1..7101

/note="wgs end extension"

clone\_end:SP6"

complement(90..201)

/rpt\_family="PB1D10"

443..539

/rpt\_family="B1F"

546..574

/rpt\_family="(TCCA)n"

complement(686..802)

/rpt\_family="PB1D9"

1155..1211

/rpt\_family="(TA)n"

1436..1618

/rpt\_family="B2\_Mm2"

2164..2185

/rpt\_family="AT\_rich"

complement(3339..3466)

/rpt\_family="B1F"

complement(4368..4448)

/rpt\_family="L1\_MM"

7102..7107

/note="clone boundary"

clone\_end:SP6

site:EcoRI"

9053..9152

/rpt\_family="(TA)n"

12029..12129

/rpt\_family="MTE"

12236..12345

/rpt\_family="PB1D7"

12752..12887

/rpt\_family="(TC)n"

complement(13407..13609)

/rpt\_family="B3"

14127..14167

```
repeat_region /rpt_family="(TA)n"
14167.14193
repeat_region /rpt_family="(CA)n"
14235.14272
repeat_region /rpt_family="(CA)n"
14521.14558
repeat_region /rpt_family="(TTA)n"
14935.14963
repeat_region /rpt_family="AT-rich"
15589.15629
repeat_region /rpt_family="(TTTTG)n"
complement(15632..15765)
repeat_region /rpt_family="B1_MM"
15919.16063
repeat_region /rpt_family="B1_MM"
16068.16090
repeat_region /rpt_family="(CAAA)n"
16143.16166
repeat_region /rpt_family="(CAAA)n"
complement(17349..17514)
repeat_region /rpt_family="RSINE1"
18604.18631
repeat_region /rpt_family="(A)n"
18696.18735
repeat_region /rpt_family="(A)n"
19606.19753
repeat_region /rpt_family="CT-rich"
22500.22598
repeat_region /rpt_family="(TTTTC)n"
22891.22945
repeat_region /rpt_family="CT-rich"
complement(24470..24595)
repeat_region /rpt_family="RSINE1"
25058.25105
repeat_region /rpt_family="(TATAA)n"
26176.26465
repeat_region /rpt_family="B4"
26473.26541
repeat_region /rpt_family="(CA)n"
26917.26939
repeat_region /rpt_family="(CAAAA)n"
28766.28877
repeat_region /rpt_family="PB1D9"
28878.28909
repeat_region /rpt_family="(CAAA)n"
29243.29455
repeat_region /rpt_family="B4"
30293.30336
repeat_region /rpt_family="AT-rich"
complement(30557..31009)
repeat_region /rpt_family="L1"
complement(31159..31213)
repeat_region /rpt_family="L1"
31219.32472
repeat_region /rpt_family="Lx4"
complement(33069..33221)
repeat_region /rpt_family="L1_MM"
33438.33470
repeat_region /rpt_family="(CAAA)n"
complement(36123..36255)
repeat_region /rpt_family="B1F"
37393.37422
repeat_region /rpt_family="AT-rich"
40393.40422
repeat_region /rpt_family="(TA)n"
40692.40712
repeat_region /rpt_family="AT-rich"
41263.41297
repeat_region /rpt_family="AT-rich"
42833.42861
repeat_region /rpt_family="AT-rich"
43372.43402
repeat_region /rpt_family="AT-rich"

complement(44537..44680)
/rpt_family="B1_MM"
complement(44781..44978)
/rpt_family="URR1A"
44993.45026
/rpt_family="AT-rich"
45864.45939
/rpt_family="CT-rich"
45948.45982
/rpt_family="(CA)n"
complement(47163..47222)
/rpt_family="5S"
47927.48187
/rpt_family="Lx8"
48211.48294
/rpt_family="Lx8"
48808.48990
/rpt_family="B3"
complement(50184..50368)
/rpt_family="B3"
50559.50602
/rpt_family="(TAA)n"

Query Match 12.0%; Score 40.2; DB 10; Length 201783;
Best Local Similarity 48.9%; Pred.No.5.3;
Matches 108; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 26 TTTTCTACAACTCAATAGATGAACAGAAATGAATGGGTAGTGACTGTTTATAAG 85
Db 87178 TTTTATAGCTAAATTTAGAAAATGCTTCATCTCTGGTATTCTTATATAATTAT 87119
QY 86 AAGAGTAATAAGATACATCATCTTTGAGCAATAAGGGAGGAGAGATTCAGCAAC 145
Db 87118 CAGACATGTTTCTTTCAGTCTAGAGGTGAGGTTATTTGAGAGGAGAGACACAGTG 87059
QY 146 AGTGTGCTTACAAGTGGAAACAACTTAAAGTCAACCCCTCTTGTACAGATCA 205
Db 87058 AGTGTGAAACAAAGTGACAGCACCATGACCTGGGCAAGGCTTCCATTATAAGAACA 86999
QY 206 ATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAAA 246
Db 86998 GGGAGAGGGAGGAGTGAAGGCGTGCCACATCATTCACA 86958

RESULT 13
CR318618/6
LOCUS 177655 bp DNA linear HTG 01-MAR-2004
DEFINITION Danio rerio clone CH211-3917, *** SEQUENCING IN PROGRESS ***, 10
unordered pieces.
ACCESSION CR318618
VERSION CR318618.1 GI:44864762
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 177655)
AUTHORS McLay, K.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zC3917
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
```

Consensus quality: 175828 bases at least Q40  
Consensus quality: 176161 bases at least Q30  
Consensus quality: 176379 bases at least Q20  
Insert size: 176755; sum-of-contigs  
Quality coverage: 188044; 8.4% error; agarose-fp  
Quality coverage: 9.57x in Q20 bases; sum-of-contigs Quality  
coverage: 9.27x in Q20 bases; agarose-fp

\*\*\*\*\*  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 57111: contig of 57111 bp in length  
\* 57112 57211: gap of 100 bp  
\* 57212 73074: contig of 15863 bp in length  
\* 73075 73174: gap of 100 bp  
\* 73175 80256: contig of 7082 bp in length  
\* 80257 80356: gap of 100 bp  
\* 80357 89660: contig of 9304 bp in length  
\* 89661 89760: gap of 100 bp  
\* 89761 98899: contig of 9139 bp in length  
\* 98900 98999: gap of 100 bp  
\* 99000 105065: contig of 6066 bp in length  
\* 105066 105165: gap of 100 bp  
\* 105166 130251: contig of 25086 bp in length  
\* 130252 130351: gap of 100 bp  
\* 130352 166939: contig of 36588 bp in length  
\* 166940 167039: gap of 100 bp  
\* 167040 172134: contig of 5095 bp in length  
\* 172135 172234: gap of 100 bp  
\* 172235 177655: contig of 5421 bp in length.

FEATURES Location/Qualifiers

1..177655  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-3917"  
/clone\_lib="CHORI-211"  
1..57111  
/note="assembly fragment:01750  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left"  
57212..73074  
/note="assembly fragment:00526  
fragment\_chain:1"  
73175..80256  
/note="assembly fragment:00319  
fragment\_chain:1"  
80357..89660  
/note="assembly fragment:00418  
fragment\_chain:2"  
89761..98899  
/note="assembly fragment:00153  
fragment\_chain:2"  
99000..105065  
/note="assembly fragment:00044  
fragment\_chain:2"  
105166..130251  
/note="assembly fragment:00847  
fragment\_chain:2"  
130352..166939  
/note="assembly fragment:01200  
fragment\_chain:2"  
167040..172134  
/note="assembly fragment:00235  
fragment\_chain:2"  
172235..177655  
/note="assembly fragment:00087

ORIGIN fragment\_chain:2

Query Match 11.8%; Score 39.8; DB 2; Length 177655;  
Best Local Similarity 52.8%; Pred. No. 6.8;  
Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 ATTATTATTTTTCATTAGATAGCCGGTTTTTTTACTTACAACTCAAATAAGATGAACAGAAATG 60  
Db 116279 ATACATTTTATTCACAGACACTCAATTTTTTTTAAATATATATAAATATGAATTAATG 116220  
Qy 61 AATGGGTTAGTACTGTTTATAAGAAGAGTAAATAAGACTACTATCATTCATTCAGGCAA 120  
Db 116219 AATTAAAGTGTAATAATAATAGAAATGTGTCAACATTTCTATATTATTATAAGTCAA 116160  
Qy 121 TAAGGGAGGAGAGATTTCAGCAACACAGTGTGCTTACAAAGTGA 163  
Db 116159 CATACAGTCTCAGCATATACAAGCACACACCCCTTACAAATGTA 116117

RESULT 14

BX537337 BX537337 217375 bp DNA linear VRT 19-APR-2004  
LOCUS Zebrafish DNA sequence from clone CH211-222E23 in linkage group 19,  
DEFINITION complete sequence.  
ACCESSION BX537337  
VERSION BX537337.9 GI:46406510  
KEYWORDS HTG  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 217375)  
AUTHORS Woodmansey,R.  
TITLE Direct Submission  
JOURNAL Submitted (17-APR-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 19, 2004 this sequence version replaced gi:46194230.

COMMENT

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfsh-help@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the rare  
assembly was confirmed by restriction digest, except on the occasion  
of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C-elegans/wormpep/clone-derived  
Zebrafish pUC subclones occasionally display inconsistency over the  
length of mononucleotide A/T runs and conserved TA repeats. Where  
this is found the longest good quality representation will be  
submitted.  
Repeat names beginning 'Dr' were identified by the Recon repeat  
discovery system (Zhirong Bao and Sean Eddy, submitted), and those  
beginning 'drr' were identified by Rick Waterman (Stephen Johnson  
Lab, WashU). For further information see





\* 87515 87614: gap of 100 bp  
\* 87615 103384: contig of 15770 bp in length  
\* 10385 103484: gap of 100 bp  
\* 10385 129543: contig of 26059 bp in length  
\* 129544 129643: gap of 100 bp  
\* 129644 162605: contig of 32962 bp in length.

FEATURES

source  
1. .162605  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14"  
/clone="RP11-555P23"  
/clone\_lib="RPCI-11 Human Male BAC"  
1. .464  
misc\_feature  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
565. .2397  
misc\_feature  
/note="assembly\_fragment"  
2498. .4504  
misc\_feature  
/note="assembly\_fragment"  
4605. .10523  
misc\_feature  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
10624. .18758  
misc\_feature  
/note="assembly\_fragment"  
18859. .27777  
misc\_feature  
/note="assembly\_fragment"  
27878. .40462  
misc\_feature  
/note="assembly\_fragment"  
40563. .53834  
misc\_feature  
/note="assembly\_fragment"  
53935. .69869  
misc\_feature  
/note="assembly\_fragment"  
69970. .87514  
misc\_feature  
/note="assembly\_fragment"  
87615. .103384  
misc\_feature  
/note="assembly\_fragment"  
103485. .129543  
misc\_feature  
/note="assembly\_fragment"  
129644. .162605  
misc\_feature  
/note="assembly\_fragment"

ORIGIN

Query Match 11.8%; Score 39.6; DB 2; Length 162605;  
Best Local Similarity 51.7%; Pred. No. 7.7;  
Matches 90; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
Qy 18 ATAGCCGGTTTTTTTACTCAACTCAATAAGATGAACAGATGAATGGTTAGTGACTGT 77  
Db 104024 ACAGTCCTTCCCCCCTACAGGTCTACAAAAGGAGGAGGAAAAGCTTTGAGTTGGATA 103965  
Qy 78 TTATAAGAAGAGTAATAAGATATCATCTTTGAGGCAATAAGGGAGGGAGAGATT 137  
Db 103964 TGAGATTGAAGTTTGTATACGAATTACTTCTTTTAGTAGTGAGTCTTAGTTT 103905  
Qy 138 CAGCAAAACAGTGTGCTTTACAGTGGAAAAACAAAGTTAAAGTGACCCCCCT 191  
Db 103904 CCATAAATAAGGCTATTATAAGTGAAAGTGTGAAAAAATGATGGGGTCTTCCT 103851

Search completed: January 14, 2005, 15:42:31  
Job time : 1499.19 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 05:22:03 ; Search time 175.111 Seconds  
(without alignments)  
10072.512 Million cell updates/sec

Title: US-10-099-663-2

Perfect score: 336

Sequence: 1 attattattttcattagata.....agacagaagaatggcattta 336

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_23Sep04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	336	100.0	336	11	ADL90128
2	336	100.0	2381	11	ADL90127 Chicken i
3	39	11.6	2933	4	ABL18616 Drosophil
4	37.2	11.1	2927	4	ABL06564 Drosophil
5	37.2	11.1	123526	10	ADJ79962 Human gli
6	36.6	10.9	403	4	AAS58239 cDNA #915
7	36.6	10.9	3322	4	ABL10846 Drosophil
8	36.6	10.9	53585	2	AAX20251 Borrelia
9	36.2	10.8	1587	4	AAH53268 S. epider
10	36.2	10.8	1878	6	ABN92204 Staphyloc
11	36.2	10.8	3000	4	AAH54543 S. epider
12	36.2	10.8	3368	4	AAH54805 S. epider
13	36.2	10.8	6088	2	AAH84331 Stealth v
14	36.2	10.8	8648	6	ABK31354 Signal tr
15	36.2	10.8	8648	6	ABL170573 Chemicali
16	35.8	10.7	714	6	ABL156240 AmEPV pho
17	35.8	10.7	50000	6	ABL56202 AmEPV gen
18	35.2	10.5	663	2	ADRO1495
19	35.2	10.5	696	2	ADRO2087
20	35.2	10.5	719	2	ADRO2446
21	35.2	10.5	856	2	ADRO1597

22	35.2	10.5	8622	6	ABL34142	Human imm
23	35.2	10.5	15714	6	ABL33173	Human imm
24	35.2	10.5	15714	6	ABQ67058	Human ang
25	35	10.4	2839	4	AAH62755	Shrimp wh
26	35	10.4	305107	4	AAH62689	Streptoco
27	34.8	10.4	18488	6	ABA01444	Streptoco
28	34.8	10.4	19738	6	ABA01436	Streptoco
29	34.8	10.4	28564	10	ADD47140	Human gen
30	34.8	10.4	75899	6	ABK85261	Human gen
31	34.8	10.4	75899	12	ADL13990	Human pro
32	34.6	10.3	6072	6	ABL32031	Human imm
33	34.4	10.2	1137	10	ADC93219	E. faeciu
34	34.4	10.2	10957	6	ABL33111	Human imm
35	34.4	10.2	12751	4	AAK85146	Human imm
36	34.4	10.2	12751	4	AAK85147	Human imm
37	34.4	10.2	69727	10	ACF65374	Phototrab
38	34.4	10.2	110000	2	AAV21209_15	Continuation (16 o
39	34.4	10.2	110000	10	ACF67367_35	Continuation (36 o
40	34.2	10.2	400	4	AAI80079	Human pol
41	34.2	10.2	7401	12	ADO26270	C35 promo
42	34.2	10.2	8605	6	ABL32535	Human imm
43	34.2	10.2	98844	7	ADJ84099	Human WRN
44	33.8	10.1	2197	4	AAD05610	Human sec
45	33.8	10.1	2664	4	AAD05581	Human sec

ALIGNMENTS

RESULT 1

ADL90128  
ID ADL90128 standard; DNA; 336 BP.

AC ADL90128;

DT 20-MAY-2004 (first entry)

DE Chicken intestinal fatty acid binding protein, iFABP, gene, promoter.  
KW Chicken; ds; intestinal fatty acid binding protein; iFABP;  
KW gut specific promoter; transgenic; promoter.  
XX Gallus gallus.

OS Gallus gallus.

PN US2003177516-A1.

XX 18-SEP-2003.

PF 14-MAR-2002; 2002US-00099663.

PR 14-MAR-2002; 2002US-00099663.

XX (HORS/) HORSEMAN N D.

PA (PRAT/) PRATT S L.

XX Horseman ND, Pratt SL;

XX WFI; 2003-898653/82.

XX New nucleic acid molecule comprising an isolated avian gut-specific gene  
expression control region, useful for regulating heterologous nucleic  
acids in transgenic avians, and for generating transgenic birds.

PS Claim 1; SEQ ID NO 2; 28pp; English.

XX The invention relates to an isolated nucleic acid comprising an isolated  
avian gut-specific gene expression control region appearing as  
CC ADL90127 (Chicken intestinal fatty acid binding protein, iFABP, gene, 5'  
region or ADL90128 (chicken iFABP promoter) or its degenerate variant.  
CC Also included are a recombinant DNA molecule comprising an isolate avian  
gut-specific gene expression control region operably linked to a nucleic  
acid insert encoding a polypeptide, an expression vector that integrates  
into a host cell (and comprising the isolated avian gut-specific gene

expression control region), expressing a heterologous polypeptide in a host cell (by transfecting a eukaryotic cell with the recombinant DNA molecule, and culturing the transfected cell in a medium suitable for expression of a heterologous polypeptide under the control of an avian intestinal fatty acid binding protein (iFABP) or cp35 gene expression control region encoded by the recombinant DNA molecule), a eukaryotic cell transformed with the expression vector (or its progeny, which expresses a heterologous polypeptide) and a transgenic avian having a heterologous polynucleotide sequence comprising the nucleic acid insert. The nucleic acids are useful for regulating heterologous nucleic acids in transgenic avians, as probes in nucleic acid hybridisation assays for detecting the iFABP gene expression control region, and for generating transgenic birds. The present sequence is the Chicken intestinal fatty acid binding protein, iFABP, gene, promoter.

XX SQ Sequence 336 BP; 116 A; 62 C; 63 G; 95 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 11; Length 336;  
Best Local Similarity 100.0%; Pred. No. 1.7e-87;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATTATTTTCATTAGATGCGGTTTTTACTCAACTCAATAAGATGAACAGATG 60  
Db |||||||  
Qy 1 ATTATTATTTTCATTAGATGCGGTTTTTACTCAACTCAATAAGATGAACAGATG 60  
Db |||||||  
Qy 61 AATGGGTTAGTACTGTTTATAAGAGAGTAATAAGATGACTATCATCTTTGGGCAA 120  
Db |||||||  
Qy 61 AATGGGTTAGTACTGTTTATAAGAGAGTAATAAGATGACTATCATCTTTGGGCAA 120  
Db |||||||  
Qy 121 TAAGGGAGGAGAGATTTCAGCAACAGTGTGCTTACAAAGTGGAACAAAGTTAAACTAAA 180  
Db |||||||  
Qy 121 TAAGGGAGGAGAGATTTCAGCAACAGTGTGCTTACAAAGTGGAACAAAGTTAAACTAAA 180  
Db |||||||  
Qy 181 GTGACCCCTCTCTTGACAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCA 240  
Db |||||||  
Qy 181 GTGACCCCTCTCTTGACAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCA 240  
Db |||||||  
Qy 241 TGTAAATTGCTTCCCTGATAGCGCTGTTCAATAATCTCTTTCGAAAGCTCTGCTACTTA 300  
Db |||||||  
Qy 241 TGTAAATTGCTTCCCTGATAGCGCTGTTCAATAATCTCTTTCGAAAGCTCTGCTACTTA 300  
Db |||||||  
Qy 301 CCAGAGTCTGCTTACAGACAGAAAGATGGCATTTA 336  
Db |||||||  
Qy 301 CCAGAGTCTGCTTACAGACAGAAAGATGGCATTTA 336  
Db |||||||

RESULT 2  
ADL90127  
ID ADL90127 standard; DNA; 2381 BP.

XX AC ADL90127;

XX DT 20-MAY-2004 (first entry)

XX DE Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region.

XX KW Chicken; ds; intestinal fatty acid binding protein; iFABP;

XX KW gut specific promoter; transgenic.

XX OS Gallus gallus.

XX PN US2003177516-A1.

XX PD 18-SEP-2003.

XX PF 14-MAR-2002; 2002US-00099663.

XX PR 14-MAR-2002; 2002US-00099663.

XX PA (HORS/) HORSEMAN N D.

XX PA (PRAT/) PRATT S L.

XX PI Horseman ND, Pratt SL;

XX DR

WPI; 2003-898653/82.

XX PT New nucleic acid molecule comprising an isolated avian gut-specific gene expression control region, useful for regulating heterologous nucleic acids in transgenic avians, and for generating transgenic birds.

XX PS Claim 1; SEQ ID NO 1; 28pp; English.

XX The invention relates to an isolated nucleic acid comprising an isolated avian gut-specific gene expression control region appearing as ADL90127 (Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region or ADL90128 (Chicken iFABP promoter) or its degenerate variant. Also included are a recombinant DNA molecule comprising an isolate avian gut-specific gene expression control region operably linked to a nucleic acid insert encoding a polypeptide, an expression vector that integrates into a host cell (and comprising the isolated avian gut-specific gene expression control region), expressing a heterologous polypeptide in a host cell (by transfecting a eukaryotic cell with the recombinant DNA molecule, and culturing the transfected cell in a medium suitable for expression of a heterologous polypeptide under the control of an avian intestinal fatty acid binding protein (iFABP) or cp35 gene expression control region encoded by the recombinant DNA molecule), a eukaryotic cell transformed with the expression vector (or its progeny, which expresses a heterologous polypeptide) and a transgenic avian having a heterologous polynucleotide sequence comprising the nucleic acid insert. The nucleic acids are useful for regulating heterologous nucleic acids in transgenic avians, as probes in nucleic acid hybridisation assays for detecting the iFABP gene expression control region, and for generating transgenic birds. The present sequence is the Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region.

XX SQ Sequence 2381 BP; 775 A; 428 C; 469 G; 709 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 11; Length 2381;  
Best Local Similarity 100.0%; Pred. No. 3.4e-87;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATTATTTTCATTAGATGCGGTTTTTACTCAACTCAATAAGATGAACAGATG 60  
Db |||||||  
Qy 1301 ATTATTATTTTCATTAGATGCGGTTTTTACTCAACTCAATAAGATGAACAGATG 1360  
Db |||||||  
Qy 61 AATGGGTTAGTACTGTTTATAAGAGAGTAATAAGATGACTATCATCTTTGGGCAA 120  
Db |||||||  
Qy 1361 AATGGGTTAGTACTGTTTATAAGAGAGTAATAAGATGACTATCATCTTTGGGCAA 1420  
Db |||||||  
Qy 121 TAAGGGAGGAGAGATTTCAGCAACAGTGTGCTTACAAAGTGGAACAAAGTTAAACTAAA 180  
Db |||||||  
Qy 1421 TAAGGGAGGAGAGATTTCAGCAACAGTGTGCTTACAAAGTGGAACAAAGTTAAACTAAA 1480  
Db |||||||  
Qy 181 GTGACCCCTCTCTTGACAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCA 240  
Db |||||||  
Qy 1481 GTGACCCCTCTCTTGACAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCA 1540  
Db |||||||  
Qy 241 TGTAAATTGCTTCCCTGATAGCGCTGTTCAATAATCTCTTTCGAAAGCTCTGCTACTTA 300  
Db |||||||  
Qy 1541 TGTAAATTGCTTCCCTGATAGCGCTGTTCAATAATCTCTTTCGAAAGCTCTGCTACTTA 1600  
Db |||||||  
Qy 301 CCAGAGTCTGCTTACAGACAGAAAGATGGCATTTA 336  
Db |||||||  
Qy 1601 CCAGAGTCTGCTTACAGACAGAAAGATGGCATTTA 1636  
Db |||||||

RESULT 3

ABL18616

ID ABL18616 standard; DNA; 2933 BP.

XX AC ABL18616;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7321.



XX PR 18-JUL-2001; 2001US-00910185.  
 XX PA (ISIS-) ISIS PHARM INC.  
 XX PI Bennett FC, Freier SM;  
 XX DR WPI; 2003-239322/23.  
 XX XX  
 XX PT New antisense oligonucleotides targeted to a nucleic acid encoding glioma  
 PT -associated oncogene-3, useful for treating developmental disorders (e.g.  
 PT holoprosencephaly) and hyperproliferative disorders (e.g. cancer).  
 XX XX  
 XX PS Disclosure; SEQ ID NO 11; 175pp; English.  
 XX CC The invention relates to a novel compound 8-50 nucleobases in length  
 CC targeted to a nucleic acid encoding glioma-associated oncogene-3 (GAO3)  
 CC or a splice variant of GAO3. The novel compound specifically hybridizes  
 CC with and inhibits the expression of GAO3 or its splice variant, or  
 CC specifically hybridizes with an 8-nucleobase portion of an active site on  
 CC a nucleic acid encoding GAO3. The antisense compound has cytostatic  
 CC activity. The antisense compound is useful for treating a disease or  
 CC condition associated with glioma-associated oncogene-3 (GAO3), such as a  
 CC developmental disorder including Greig's cephalopolysyndactyly, Pallister  
 CC -Hall syndrome, post-axial polydactyly, holoprosencephaly, Rubenstein-  
 CC Teybi syndrome or basal cell nevoid syndrome, and a hyperproliferative  
 CC disorder, such as cancer. This polynucleotide represents a DNA sequence  
 CC relating to the human glioma-associated oncogene-3 (GAO3) of the  
 CC invention.  
 XX SQ Sequence 123526 BP; 34368 A; 24882 C; 26026 G; 38250 T; 0 U; 0 Other;  
 Query Match 11.1%; Score 37.2; DB 10; Length 123526;  
 Best Local Similarity 53.4%; Pred. No. 8;  
 Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 QY 12 CATTAGATACCGGTTTTTTTACTACAACTCAAATAAGATGAACAGAAATGAATGGGTAGT 71  
 Db 107264 CATTCCCAATGCCCTTTTAATTAATGCTACAGAAATGCACAGAGAGAGGGGTGCT 107205  
 QY 72 GACTGTTTAAAGAGAGAGTAATAAGATATCTATCATCTTTCAGGCAATAAGGGAGGGA 131  
 Db 107204 GTGTACCCATTAAATAATGAGATGAAGAAAGAAAGAAAGGCAAGAGGGAGGGA 107145  
 QY 132 GAGATTACAGCAACAGTGCTTACA 157  
 Db 107144 GGGAGAAACCGAAAGCATCCATGCA 107119  
 RESULT 6  
 AAS58239  
 ID AAS58239 standard; cDNA; 403 BP.  
 XX AC AAS58239;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE cDNA #915 encoding portion of a human colon tumour protein.  
 XX KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
 XX OS Homo sapiens.  
 XX PN WO200173027-A2.  
 XX PD 04-OCT-2001.  
 XX PF 22-MAR-2001; 2001WO-US009246.  
 XX PR 24-MAR-2000; 2000US-0191597P.  
 XX PI 04-MAY-2000; 2000US-0202024P.  
 XX DR 05-MAY-2000; 2000US-0202189P.

PA (CORI-) CORIXA CORP.  
 XX Meagher MJ, Xu J, King GE;  
 XX WPI; 2001-611627/70.  
 XX XX  
 XX PT New colon tumor proteins and related nucleic acid, useful for treatment,  
 PT prevention, diagnosis and monitoring of cancer.  
 XX XX  
 XX PS Claim 4; Page 202; 299pp; English.  
 XX CC Th present invention relates to the isolation of novel cDNA sequences  
 CC encoding for at least an immunogenic portion of human colon tumour  
 CC proteins. The sequences of the invention are useful in pharmaceutical  
 CC compositions and vaccines for the prevention and treatment of cancers  
 CC such as colon cancer. They are also useful for the diagnosis and  
 CC monitoring of such cancers. Antibodies to the colon tumour proteins and  
 CC antigen presenting cells that express polynucleotides encoding colon  
 CC tumour proteins can be used to inhibit the development of cancers. T-  
 CC cells that react specifically with colon tumour proteins are useful for  
 CC removing tumour cells from samples (e.g. blood) and for cancer treatment.  
 CC The polynucleotides sequences are also useful in gene therapy. AAS57325-  
 CC AAS58880 represent the cDNA sequences of the invention that encode for  
 CC portions of human colon tumour proteins  
 XX SQ Sequence 403 BP; 136 A; 63 C; 88 G; 105 T; 0 U; 11 Other;  
 Query Match 10.9%; Score 36.6; DB 4; Length 403;  
 Best Local Similarity 50.6%; Pred. No. 1.7;  
 Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
 QY 26 TTTTITACTACAACTCAATATAGATGAACAGATGAATGGGTAGTACTGTTATAAAG 85  
 Db 178 TCTTTAAAGTTTAAAGAAAAAGAGCTGCAGAGTATTTATAAACTGCTCTTTAGAAAAA 237  
 QY 86 AAGAGTAAATAAGATCTATCATCTTTGAGGCAATTAAGGGAGGAGAGATTTCAGCAAC 145  
 Db 238 AACAGCAGAGAGACCATTTGACCATATGATGGAAGAGGGAAGAGTATTATAGAAC 297  
 QY 146 AGTGTCTTACAGTGGAAAAACAAGTTAAACTTAAAGTACCCCTT 191  
 Db 298 TTTGCTAGTTNAAAAAATAAAAAAATAAACTTNGCNGAACCCCT 343  
 RESULT 7  
 ABL10846  
 ID ABL10846 standard; cDNA; 3322 BP.  
 XX AC ABL10846;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27020.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical; gene; ss.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX XX  
 XX PA (PEKE ) PE CORP NY.  
 XX XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.

DR P-PSDB; ABB66743.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from *Drosophila* and for elucidating cell signaling and cell-cell

PT interactions.

XX Claim 1; SEQ ID NO 27020; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-

CC ABB572072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 3322 BP; 993 A; 715 C; 742 G; 872 T; 0 U; 0 Other;

Query Match 10.9%; Score 36.6; DB 4; Length 3322;

Best Local Similarity 57.4%; Pred. No. 3.5;

Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 ATTATTATTTTATTAGATAGCCGGTTTCTTACTCAAACTCAATAAGATGAACAGAAATG 60

Db 3191 ATTGAGATTCCTTTTATCTGTTTGTATTATCTACAAATTTGGTGCCAGTAGA 3250

Qy 61 AATGGGTTAGTACTGTTTATAAGAGAGAGTAATAAGATGACTATCATCATTTGA 115

Db 3251 AGTTAGATTTTAATCGGTTAGCAGGAAGTGCAATTAACAAATTATTATACTTGA 3305

RESULT 8

AAAX20251/c

ID AAX20251 standard; DNA; 53585 BP.

XX

AC AAX20251;

XX

DT 04-MAY-1999 (first entry)

XX

XX *Borrelia burgdorferi* polynucleotide sequence #4.

DE

XX *Borrelia burgdorferi*; spirochete; bacterium; pathogen; Lyme disease;

KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

KW infection; diagnosis; characterisation; detection; ds.

XX

OS *Borrelia burgdorferi*.

XX

XX WO9858943-A1.

XX

XX 30-DEC-1998.

XX

XX 18-JUN-1998; 98WO-US012764.

XX

XX 20-JUN-1997; 97US-0050359P.

PR 22-JUL-1997; 97US-0053344P.

PR 22-JUL-1997; 97US-0053377P.

PR 03-SEP-1997; 97US-0057483P.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA (MED-) MEDIMMUNE INC.

XX

XX Fraser C, White OR, Clayton R, Dougherty BA, Lathigra R;

PI Smith HO;

XX

XX WPI; 1999-081217/07.

XX

XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention and

PT therapy of infections, particularly Lyme disease.

XX

PS Claim 1; Page 801-831; 1128pp; English.

XX

CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from

CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for the

CC detection, diagnosis, characterisation, prevention and therapy of Bb

CC infections, e.g. Lyme disease. They can also be used for the production

CC of biosynthetic products, e.g. enzymes. *Borrelia* belongs to a family of

CC motile, spiral-shaped bacteria called Spirochetes. Spirochetes are

CC pathogenic in humans and *Borrelia* causes epidemic and endemic relapsing

CC fever, and Lyme borreliosis, more commonly known as Lyme disease

XX

SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T; 0 U; 2 Other;

Query Match 10.9%; Score 36.6; DB 2; Length 53585;

Best Local Similarity 58.9%; Pred. No. 9;

Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 78 TTATAAGAGAGAGTAATAAGATGACTATCATCTTTCAGGCAATAAGGGAGGAGATT 137

Db 44978 TTAATAATGTTAATAATAAGATTACTCAACTTATATCAAGCATTAATGATGAGAGATG 44919

Qy 138 CAGCAACACAGTGTCTTACAGTGGAACAAAGTTAACTAAAGTGA 184

Db 44918 CAAGCATCATTTGAACATTTTAAGAGAGAGAGAAATGAACATAATGTCA 44872

RESULT 9

AAHS3268

ID AAHS3268 standard; DNA; 1587 BP.

XX

AC AAHS3268;

XX

DT 03-SEP-2001 (first entry)

XX

XX *S. epidermidis* open reading frame nucleotide sequence SEQ ID NO:1929.

DE

XX *Staphylococcus epidermidis* SRI strain; infection; diagnosis; vaccination;

KW endocarditis; ds.

XX

OS *Staphylococcus epidermidis*.

XX

XX WO200134809-A2.

XX

XX 17-MAY-2001.

XX

XX 09-NOV-2000; 2000WO-US030782.

PF

XX 09-NOV-1999; 99US-0164258P.

PR

XX (GLAX ) GLAXO GROUP LTD.

XX

XX Kimmerly WJ;

XX

XX WPI; 2001-316495/33.

DR

XX P-PSDB; AAG82418.

XX

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,

PT useful for vaccinating against infections, e.g. endocarditis.

XX

XX Claim 8; Page 527-528; 2188pp; English.

XX

XX AAHS52304 to AAHS53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I)

CC and (II) can have antibacterial activity and therefore can be used in

CC vaccination. The nucleic acids (I) may be used to produce the S.

CC *epidermidis* polypeptides (II) via the production of vectors containing

CC them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAHS53971 to

CC AAHS5090 represent specifically claimed *S. epidermidis* genomic DNA

CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464

XX  
SQ Sequence 1587 BP; 617 A; 268 C; 201 G; 501 T; 0 U; 0 Other;  
Query Match 10.8%; Score 36.2; DB 4; Length 1587;  
Best Local Similarity 53.1%; Pred. No. 3.5;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTACAAAGTGGAAACAAGTTAAACTAAAGTACCCCTCTCTGACAA 200  
DB 203 CAGACGATGACTTTAAAGTTGAAACCTTTATTTTACAAGAGTTACCTCAAGCCCATCTA 262  
QY 201 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATGTAAATTTGCTTTCTTGATA 260  
DB 263 GAAATAATAACAATATCTCAAGCATATAGATAATCAATATCAATCTTCTGAAA 322  
QY 261 AGCCTGTTTCATAAATCTCTTTGCA 285  
DB 323 ATAAATACAACTTACTTGCA 347

RESULT 10  
ABN92204  
ID ABN92204 standard; DNA; 1878 BP.  
XX  
AC ABN92204;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1667.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.  
XX US6380370-B1.  
XX 30-APR-2002.  
XX 13-AUG-1998; 98US-00134001.  
XX 14-AUG-1997; 97US-0055779P.  
XX 08-NOV-1997; 97US-0064964P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2002-381255/41.  
XX P-PSDB; ABP39659.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
XX polypeptide, useful for diagnosing and treating bacterial infections.  
XX  
XX Disclosure; SEQ ID NO 1667; 267pp; English.  
XX

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences  
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have  
XX antibacterial activity and can be used in gene therapy. The sequences can  
XX also be used in the diagnosis and treatment of bacterial infections,  
XX particularly S. epidermidis infections. The sequences can be used to  
XX screen for compounds able to interfere with the S. epidermidis life cycle  
XX or inhibit S. epidermidis infection. N.B. The sequence data for this  
XX patent did not form part of the printed specification, but was obtained

CC in electronic format directly from the USPTO web site  
XX  
SQ Sequence 1878 BP; 734 A; 304 C; 239 G; 601 T; 0 U; 0 Other;  
Query Match 10.8%; Score 36.2; DB 6; Length 1878;  
Best Local Similarity 53.1%; Pred. No. 3.7;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTACAAAGTGGAAACAAGTTAAACTAAAGTACCCCTCTCTGACAA 200  
DB 494 CAGACGATGACTTTAAAGTTGAAACCTTTATTTTACAAGAGTTACCTCAAGCCCATCTA 553  
QY 201 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATGTAAATTTGCTTTCTTGATA 260  
DB 554 GAAATAATAACAATATCTCAAGCATATAGATAATCAATATCAATCTTCTGAAA 613  
QY 261 AGCCTGTTTCATAAATCTCTTTGCA 285  
DB 614 ATAAATACAACTTACTTGCA 638

RESULT 11  
AAH54543  
ID AAH54543 standard; DNA; 3000 BP.  
XX  
AC AAH54543;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3907.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;  
KW endocarditis; ds.

XX Staphylococcus epidermidis.  
XX WO200134809-A2.  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US030782.  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis.  
XX  
XX Claim 8; Page 1560-1561; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
XX and (II) can have antibacterial activity and therefore can be used in  
XX vaccination. The nucleic acids (I) may be used to produce the S.  
XX epidermidis polypeptides (II) via the production of vectors containing  
XX them which are used to produce host cells which express the  
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
XX used to vaccinate subjects and to raise antibodies against the bacteria.  
XX The polypeptides may also be used to assay for other inhibitors of their  
XX activity and therefore identify compounds that may be used for the  
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA  
XX polynucleotide sequences from the present invention. AAH5091 to AAH5098  
XX represent oligonucleotide sequences and primers which are used in the  
XX exemplification of the present invention. N.B. The present invention  
XX specifically claims all the polynucleotide sequences given in the  
XX sequence listing of the present specification, however the sequence  
XX listing only goes up to SEQ ID NO:4454 so even though sequences are given



CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 3000 BP; 1019 A; 503 C; 512 G; 966 T; 0 U; 0 Other;  
Query Match 10.8%; Score 36.2; DB 4; Length 3000;  
Best Local Similarity 53.1%; Pred. No. 4.4;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 141 CAACACGTGCTTACAGTGGAACAAGTTAACTAAAGTGACCCCTCCTTGACAA 200  
Db 129 CAGACGATGACTTTAAAGTTGAAAACCTTATTTTACAAGAGTTACCTCAAGCCCATACTA 188  
QY 201 GATCAATGCCACAGTTCAGCTTTAGCCAGCCACATCATGTAAATGCTTTCTCGATA 260  
Db 189 GAAAAATAAACAAATATCTCAAGCATATAGATATCATCAATTATCAACTTCGAAA 248  
QY 261 AGCCTGTTCATAAATCTCTTTGCA 285  
Db 249 ATAAATACAACAATTACTTGTGCA 273  
RESULT 12  
AAH54805/c  
ID AAH54805 standard; DNA; 3368 BP.  
XX AC AAH54805;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4169.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;  
KW endocarditis; ds.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US030782.  
XX  
PR 09-NOV-1999; 99US-0164258P.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
PI Kimmerly WJ;  
XX  
DR WPI; 2001-316495/33.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
PS Claim 8; Page 1875-1876; 2189pp; English.  
XX  
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence

CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 3368 BP; 1163 A; 447 C; 569 G; 1189 T; 0 U; 0 Other;  
Query Match 10.8%; Score 36.2; DB 4; Length 3368;  
Best Local Similarity 53.1%; Pred. No. 4.5;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 141 CAACACGTGCTTACAGTGGAACAAGTTAACTAAAGTGACCCCTCCTTGACAA 200  
Db 1038 CAGACGATGACTTTAAAGTTGAAAACCTTATTTTACAAGAGTTACCTCAAGCCCATACTA 979  
QY 201 GATCAATGCCACAGTTCAGCTTTAGCCAGCCACATCATGTAAATGCTTTCTCGATA 260  
Db 978 GAAAAATAAACAAATATCTCAAGCATATAGATATCATCAATTATCAACTTCGAAA 919  
QY 261 AGCCTGTTCATAAATCTCTTTGCA 285  
Db 918 ATAAATACAACAATTACTTGTGCA 894  
RESULT 13  
AAH84331  
ID AAX84331 standard; DNA; 6088 BP.  
XX AC AAX84331;  
XX  
DT 08-SEP-1999 (first entry)  
XX  
DE Stealth virus nucleic acid clone, SEQ ID NO: 23.  
XX  
KW Stealth virus; detection; diagnosis; infection; ss.  
XX  
OS Stealth virus.  
XX  
PN Key Location/Qualifiers  
FT misc\_difference 5841  
FT /tag= a  
FT /note= "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the base  
FT numbering given in the specification"  
FT misc\_feature 5997..8133  
FT /tag= b  
FT /note= "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the base  
FT numbering given in the specification"  
FT misc\_difference 8275  
FT /tag= d  
FT /note= "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the base  
FT numbering given in the specification"  
XX  
PN WO9934019-A1.  
XX  
PD 08-JUL-1999.  
XX  
PF 30-DEC-1998; 98WO-US027744.  
XX  
PR 30-DEC-1997; 97US-00001184.  
XX  
PA (MART/) MARTIN W J.  
XX  
PI Martin WJ;  
XX  
DR WPI; 1999-405521/34.  
XX  
PT Novel strains of stealth virus.  
XX  
PS Claim 19; Page 64-66; 95pp; English.  
XX  
CC This sequence represents a Stealth virus nucleic acid clone. The

invention relates to a method of detecting and characterising a stealth virus by reacting a sample suspected of containing a stealth virus with a probe from a known stealth virus and sequencing the resultant isolated nucleotide. The method comprises the steps of: (a) isolating DNA or RNA from a sample suspected of containing a stealth virus, e.g. a culture of cells showing a viral cytopathic effect; (b) testing the reactivity of the isolated DNA or RNA with a molecular probe that contains at least 18 or more contiguous nucleotides identical to sequence previously identified from a stealth virus; and, optionally (c) sequencing the isolated DNA or RNA molecules that react with the probe. The method is used to detect stealth virus in a biological product, food or in the environment. The method is also used to evaluate agents for their inhibitory or stimulatory effects on stealth virus replication and to determine capacity of the virus to recombine with and potentially alter the nucleic acid sequences of a cell or bacterium

Query Match	10.8%	Score	36.2	DB	2	Length	6088
Best Local Similarity	20.1%	Pred. No.	5.6				
Matches	43	Conservative	77	Mismatches	94	Indels	0
						Gaps	0

Qy	42	AAATAAGTGAACAGAAATGAATCGGTTASTGACTGTTTATATAAGAAGAGAGTAATAAAGATA	101
Db	5824	AAATTGAAGAAATTCWHANYVWTSNNRMAMHADATAPAGRAHNTRMHNMAITNDATNST	5883

QY 102 CTATCATCTTGAGGCAATAAGGGAGGGAGAGATTTCAGCAACACTGTGCTTACAAGTG 161

D**b** 5884 A THBCDCMBGTMSNRWMSYMBARAHBCWHNMARTNWHNMARTNHHDDTBCBWHNMARTN 5943

162 GAAAACAAGTTAAACTAAAGTGAACCCCTCTTGACAAGATCAATGCCACAGTTGAGCT 221

DD NKMTAWHNWAKINIMCKRSIWKDMMDDHCCDBBCTIDGDANBCDAAADKINIKRYTABNWRJDCM 6000

[illegible]

RESULT 14  
ABK31354

[illegible]

XX  
DT 23-APR-2002 (first entry)

DE Signal transduction associated gene modified DNA #99.

Human; signal transduction associated gene; cytosine methylation state;

antitumour; cytostatic; mutant; ds.

OS Homo sapiens.  
OS Synthetic.

XX PN WO200200926-A2.

03 - JAN - 2002 .  
PD  
vuz

PF 29-JUN-2001; 2001WO-EP007472.  
XX

PR 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;  
vv

DR WPI; 2002-147896/19.  
XX

PT Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic

PT sequences of genes associated with signal transduction.  
XX  
PS Claim 1; SEQ ID NO 197; 24pp; English.

The present invention relates to chemically modified DNA sequences of signal transduction associated genes. The DNA sequences are chemically modified using a solution of bisulphite, hydrogen sulphite or disulphite. Also disclosed are oligonucleotides and/or DNA oligomers for detecting the cytosine methylation state (CpG islands) of these genes, and a method for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with signal transduction. The genomic DNA can be obtained from cells or cellular components which contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as tissue from eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver, histologic object slides, and all their possible combinations. The sequences of the invention are useful for the diagnosis and therapy of diseases associated with signal transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA sequences of different genes associated with signal transduction, or their complementary sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

Sequence 8648 BP: 2149 A; 236 C; 1917 G; 4346 T; 0 U; 0 Other;

Query Match	10.8%	Score 36.2;	DB 6;	Length 8648;
Best Local Similarity	57.5%;	Pred. No. 6.3;		
Matches 65;	Conservative	0;	Mismatches 48;	Indels 0;
				Gaps 0;

3 TATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAATAAGATGAACACAGATGAA 62

Db 940 TTTTGGTTTTATTACGTGGTTAGTTTATTAGTAGATTTATGATAAGTTTAGTTTTCATAT 999

QY 63 TGGGTTAGTGACTGTTTATAAAGAAGAGCTAAATAAGATACATCATCATTTGA 115

DD 1000 11GGGAGTGAATGTTTTATGTGAGAGTAAATAGATTATATTTTGGTTATTGGA 1052

RESULT 15  
ADP 70572

ID ABL70573 standard; DNA; 8648 BP.  
XX

AC ABL70573;  
XX XX[illegible]

Cell signalling: cytosine methylation: cell signalling disease: cancer.

KW tumour; cytostatic; ds.  
XX

OS  
XX  
unidentified.

XX  
XX  
10 MAY 2005

XX PF 29-JUN-2001: 2001WO-EP007471.

PR 30-JUN-2000; 2000DE-01032529.

01-SEP-2000; 2000DE-01073020.  
XX  
XX  
/ 0000 / 0000000000 00

XX  
BT    Olek A Bienenbrock C Ber]in V.  
XX

XX  
DR  
WPI: 2002-154758/20.

XX  
PT Nucleic acid, useful for diagnosis and therapy of diseases associated

PT sequences of genes associated with cell signaling.

Search completed: January 14, 2005, 09:20:25  
Job time : 182.111 secs

**THIS PAGE IS BLANK**

Result No.	Score	Query Match	Length	DB	ID	Description
1	37.2	11.1	1141	4	US-09-806-708B-22	Sequence 22, Appl
2	36.2	10.8	1587	4	US-09-710-279-1929	Sequence 1929, Ap
3	36.2	10.8	1878	3	US-09-134-001C-1667	Sequence 1667, Ap
4	36.2	10.8	3000	4	US-09-710-279-3907	Sequence 3907, Ap
C 5	36.2	10.8	3368	4	US-09-710-279-4169	Sequence 4169, Ap
C 6	35.8	10.7	7218	1	US-08-232-463-14	Sequence 14, Appl
C 7	35.2	10.5	663	3	US-08-998-416-187	Sequence 187, App
C 8	35.2	10.5	696	3	US-08-998-416-779	Sequence 779, App
C 9	35.2	10.5	719	3	US-08-998-416-1138	Sequence 1138, Ap
C 10	35.2	10.5	856	3	US-08-998-416-289	Sequence 289, App
11	35	10.4	399	4	US-09-621-976-8976	Sequence 8976, Ap
C 12	34.4	10.2	1137	4	US-09-107-532A-2846	Sequence 2846, Ap
C 13	34.4	10.2	1664976	4	US-08-915-421B-1	Sequence 1, Appl
C 14	34.4	10.2	1664976	4	US-09-692-570-1	Sequence 10, Appl
C 15	34.2	10.2	98844	4	US-09-791-211-10	Sequence 7, Appl
C 16	33.6	10.0	2119	3	US-09-240-639-7	Sequence 7, Appl
C 17	33.6	10.0	2119	4	US-09-308-510A-7	Sequence 7, Appl
C 18	33.6	10.0	2119	4	US-09-905-744-7	Sequence 7, Appl
C 19	33.6	10.0	2119	4	US-10-107-660-7	Sequence 7, Appl
C 20	33.6	10.0	2119	4	US-10-107-576-7	Sequence 7, Appl
C 21	33.6	10.0	2119	4	US-09-905-732-7	Sequence 7, Appl
22	33	9.8	2642	1	US-08-178-242-4	Sequence 4, Appl
23	33	9.8	2642	2	US-08-955-091-4	Sequence 4, Appl
24	33	9.8	2642	3	US-09-225-510-4	Sequence 4, Appl
25	33	9.8	2870	1	US-08-178-242-14	Sequence 14, Appl
26	33	9.8	2870	2	US-08-955-091-14	Sequence 14, Appl
27	33	9.8	2870	3	US-09-225-510-14	Sequence 14, Appl

```
US-09-710-279-1929
; Sequence 1929, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1929
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1929

Query Match          10.8%; Score 36.2; DB 4; Length 1587;
Best Local Similarity 53.1%; Pred. No. 0.24;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTACAGTGGAAACAAAGTTAAACTAAAGTGACCCCTCTCTTGACAA 200
Db 203 CAGACGATGACTTTAAAGTTGAAACCTTTATTTTACAAGAGTTACCTCAAGCCCATACTA 262
QY 201 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATGTTCTTCCTGATA 260
Db 263 GAAATAATAACAATAATATCTCAAGCATATAGATATATCAATATCAATCTCTGAAA 322
QY 261 AGCCTGTTCTATAAATCTCTTTGCA 285
Db 323 ATAAATACAACAATTACTTTGTGCA 347

RESULT 3
US-09-134-001C-1667
; Sequence 1667, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1667
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1667

Query Match          10.8%; Score 36.2; DB 3; Length 1878;
Best Local Similarity 53.1%; Pred. No. 0.25;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTACAGTGGAAACAAAGTTAAACTAAAGTGACCCCTCTCTTGACAA 200
Db 494 CAGACGATGACTTTAAAGTTGAAACCTTTATTTTACAAGAGTTACCTCAAGCCCATACTA 553
QY 201 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATGTTCTTCCTGATA 260
Db 554 GAAATAATAACAATAATATCTCAAGCATATAGATATATCAATATCAATCTCTGAAA 613
```

```
QY 261 AGCCTGTTCTATAAATCTCTTTGCA 285
Db 614 ATAAATACAACAATTACTTTGTGCA 638

RESULT 4
US-09-710-279-3907
; Sequence 3907, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3907
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3907

Query Match          10.8%; Score 36.2; DB 4; Length 3000;
Best Local Similarity 53.1%; Pred. No. 0.3;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTCAAGTGGAACAAAGTTAAACTAAAGTGACCCCTCTCTTGACAA 200
Db 129 CAGACGATGACTTTAAAGTTGAAACCTTTATTTTACAAGAGTTACCTCAAGCCCATACTA 188
QY 201 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATGTTCTTCCTGATA 260
Db 189 GAAATAATAACAATAATATCTCAAGCATATAGATATATCAATATCAATCTCTGAAA 248
QY 261 AGCCTGTTCTATAAATCTCTTTGCA 285
Db 249 ATAAATACAACAATTACTTTGTGCA 273

RESULT 5
US-09-710-279-4169/c
; Sequence 4169, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4169
; LENGTH: 3368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4169

Query Match          10.8%; Score 36.2; DB 4; Length 3368;
Best Local Similarity 53.1%; Pred. No. 0.32;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTCAAGTGGAACAAAGTTAAACTAAAGTGACCCCTCTCTTGACAA 200
```



```
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippesen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 779:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1501RP
; US-08-998-416-779

Query Match 10.5%; Score 35.2; DB 3; Length 696;
Best Local Similarity 58.7%; Pred. No. 0.35;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 ATTATTATTTTCATTAGATAGCGGGTTTTTACTACAACTCAAAATGAAGTGAACAGAAATG 60
Db 460 ATTATTAACCTTTATTAGTTAAACCAITTTATTATTAATGATCATAAATAATAAAGGAATAA 519
QY 61 AATGGGTTAGTACTGTTTATAAAGAGAGCTAATAAAGATACTA 104
Db 520 CATTAATGATATATAGTTATTATTAAGAACCAATGAAGATACTA 563

RESULT 9
US-08-998-416-1138/c
; Sequence 1138, Application US/089998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippesen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
```

```
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 719 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692UP
; US-08-998-416-1138

Query Match 10.5%; Score 35.2; DB 3; Length 719;
Best Local Similarity 58.7%; Pred. No. 0.35;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 ATTATTATTTTCATTAGATAGCGGGTTTTTACTACAACTCAAAATGAAGTGAACAGAAATG 60
Db 567 ATTATTAACCTTTATTAGTTAAACCAITTTATTATTAATGATCATAAATAATAAAGGAATAA 508
QY 61 AATGGGTTAGTACTGTTTATAAAGAGAGCTAATAAAGATACTA 104
Db 507 CATTAATGATATATAGTTATTATTAAGAACCAATGAAGATACTA 464

RESULT 10
US-08-998-416-289/c
; Sequence 289, Application US/089998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippesen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
```



```
/ STATE: No. 6239264th Carolina
/ COUNTRY: USA
/ ZIP: 27709
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/998,416
/ FILING DATE: 24-DEC-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: CH 0016/97
/ FILING DATE: 31-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weigs, J. Timothy
/ REGISTRATION NUMBER: 38,241
/ REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 919-541-8587
/ TELEFAX: 919-541-8689
/ INFORMATION FOR SEQ ID NO: 289:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 856 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: PAG1241UP
/ US-08-998-416-289

Query Match 10.5%; Score 35.2; DB 3; Length 856;
Best Local Similarity 58.7%; Pred. No. 0.38; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 43;

QY 1 ATTATTATTTTCATTAGTAGCGGGTTTTTTTACTACAACTCAATAAGATGAACAGAATG 60
DB 567 ATTATTAACTTTATTAGTTAAACCATTTTAAATGATCATTAATATATTAAGGAATAA 508

QY 61 AATGGGTTAGTGACTGTTTATATAAGAAGAGTAATAAGATACTA 104
DB 507 CATTAATCATATATAGTTATTATATAGAACCAATGAAGATACTA 464

RESULT 11
US-09-621-976-8976
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 10.4%; Score 35; DB 4; Length 399;
Best Local Similarity 11.7%; Pred. No. 0.32;
Matches 29; Conservative 114; Mismatches 104; Indels 0; Gaps 0;

QY 50 TGAACAGAAATGAATGGGTAGTACTGTTTATAAGAAGAGTAATAAGATACTATCATC 109
DB 43 KRRRRRRAMWVKSCMWMKSKSWRSWGMWTKRMKGRGAASWAGYMSWMTYMTTRWRY 102

QY 110 ATTTGAGGCAATAAGGAGGAGAGATTTCAGCAAAACAGTGTGCTTTACAAAGTGAACAA 169
DB 103 RYRKCACTKWAAGAGWKGAGWAWAYAKWYMAWRTAMKYWAMMKSKSRMRRAWYAW 162

QY 170 GTTAAACTAAAGTGACCCCCCTCTTGACAAGATCAATGCGCACAGTTGAGCTTTAGCCAG 229
DB 163 MYVMARRTMWMGRASCYRGAYMASAGWYMYMYMMRRKWWYSAGWSMKRKTWTRCASY 222

QY 230 CCACATCATCATGTAATGCTTTCCCTCGATAAGCTGTTTCATAAATCTCTTTGCAAGC 289
DB 223 SCWSSYCMWGAAMWYWKTSRWSYSSYRCKTYRRSCCCWMSYWKTYRSMYWCASY 282

QY 290 TCTGCTA 296
DB 283 YSYKTKR 289

RESULT 12
US-09-107-532A-2846
; Sequence 2846, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2846:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1137
; SEQUENCE DESCRIPTION: SEQ ID NO: 2846:
US-09-107-532A-2846

Query Match 10.2%; Score 34.4; DB 4; Length 1137;
Best Local Similarity 52.0%; Pred. No. 0.73;
```

```

Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 1 ATTATTATTTTATTAGATGACCGGTTTTTACTACAACTCAAAATAGATGAACAGATG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 ATCATAACTATGAAGAGATCAACGTTTGATTGAAACGTTTGAAGAAGATGAGATGAGA 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 AATGGGTAGTACTCTTTTATRAAGAGAGTATATAAGATACTATCATCATTTTGAGGCAA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 AAGAATTTGGGAAAGTTTAGAGAAAGCAACAACTGTCGAAGCAATCTTTACATGCAA 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 TAAGGGAGGAGAGATTTCAGCAACACAGT 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 AARACACGGGATGTCACACCAATGGT 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
;
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)

```

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

```
Query Match 10.2%; Score 34.4; DB 4; Length 1664976;
Best Local Similarity 49.4%; Pred. No. 12;
Matches 89; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 34 TACAACCTCAATAGATGAACAGAAATGAATGGGTTAGTACTGTTTATAAGAAGAGTAA 93
Db 1555109 TAAAGTTAAAGTTATCTACAAATGGAATAAATCCATGGGAATTCATATATTAATTTAAGTTG 1555050

Qy 94 TAAGATACATCATCTATTTGAGCAATAAGGGGAGAGATTCACCAACAGTGCT 153
Db 1555049 GGAGGAGAAATAAACCTTTAGAAGGATATAGGATTCAGATGATGAATAATGATTTT 1554990

Qy 154 TACAAGTGGAAAAAAGTTAACTAAAGTGACCCCTTCCTTGACAAGATCAATGCCACA 213
Db 1554989 ATTTGTTGAGATTAACATATCAAAAGGNAATAGATTAATTAATAGAGCAATGCCAAA 1554930
```

```
RESULT 14
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
```

```
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10398)..(10398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
```



; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 98844
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 24962
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 64383
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65468
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65469
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65470
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65471
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 87130
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 89049
; OTHER INFORMATION: unknown
; OTHER INFORMATION:
;
US-09-791-211-10

Query Match 10.2%; Score 34.2; DB 4; Length 98844;
Best Local Similarity 58.3%; Pred. No. 4.7;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 45 TAAGATGAACAGATGAATGGTTAGTGACTGTTTATAAAGAGAGTAATAAAGATACTA 104
Db 34669 TAGCATGAACCAATGCATTTTAACGTTAATAATGAGGAAAGGGAATGCATGATAATT 34728

Qy 105 TCATCATTTGAGGCAATAAGGGAGGAGATTTCAGCAACAG 147
Db 34729 TCATCAAGTCAGAAAAAATGGAAGAAAAAGAGAGAACTG 34771

Search completed: January 14, 2005, 15:46:46
Job time : 42.2874 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2005, 09:20:32 ; Search time 196.381 Seconds  
(without alignments)  
9830.969 Million cell updates/sec

Title: US-10-099-663-2  
Perfect score: 336  
Sequence: 1 attattatttcattagata.....agacagaagatggcattta 336

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues  
Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/PCT PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06 NEW PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06 PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07 NEW PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08 NEW PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08 PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09A PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09B PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09C PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US10A PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10B PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10C PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10D PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10E PUBCOMB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq:\*

20: /cgn2\_6/ptodata/2/pubpna/US60 NEW PUB.seq:\*

21: /cgn2\_6/ptodata/2/pubpna/US60 PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	336	100.0	336	15	US-10-099-663-2
2	336	100.0	2381	15	Sequence 2, Appli
3	40.4	12.0	116327	18	Sequence 1, Appli
c	37.2	11.1	123526	10	Sequence 6867, Ap
	36.6	10.9	403	9	Sequence 11, Appl
	36.6	10.9	403	16	Sequence 915, App
	36.2	10.8	8648	16	Sequence 915, App
	36.2	10.8	8648	16	Sequence 217, App
8	35.4	10.5	3673778	15	Sequence 1, Appli
c	35.2	10.5	479	16	Sequence 47091, A
c	35.2	10.5	479	16	Sequence 47091, A
11	35.2	10.5	8622	15	Sequence 2115, Ap
c	35.2	10.5	15714	15	Sequence 1146, Ap

c	13	35.2	10.5	15714	17	US-10-433-793-88	Sequence 88, Appl
c	14	35	10.4	134738	13	US-10-087-192-502	Sequence 502, App
c	15	34.8	10.4	75899	9	US-09-854-883-243	Sequence 243, App
c	16	34.8	10.4	75899	15	US-10-360-510-243	Sequence 243, App
	17	34.6	10.3	6072	15	US-10-311-455-4	Sequence 4, Appli
	18	34.4	10.2	1173	15	US-10-369-493-24780	Sequence 24780, A
	19	34.4	10.2	10957	15	US-10-311-455-1084	Sequence 1084, A
	20	34.2	10.2	598	17	US-10-767-701-24993	Sequence 24993, A
c	21	34.2	10.2	8605	15	US-10-311-455-508	Sequence 508, App
	22	34.2	10.2	151870	17	US-10-741-601-5614	Sequence 5614, Ap
c	23	34	10.1	290547	17	US-10-367-094-77	Sequence 77, Appl
	24	33.8	10.1	1028	17	US-10-767-795-4283	Sequence 4283, Ap
	25	33.8	10.1	1200	9	US-09-887-576-755	Sequence 755, App
c	26	33.8	10.1	2542	13	US-10-027-632-111567	Sequence 111567,
c	27	33.8	10.1	2542	15	US-10-027-632-111567	Sequence 111567,
c	28	33.8	10.1	4127	16	US-10-374-780A-2169	Sequence 2169, Ap
	29	33.8	10.1	127238	13	US-10-087-192-787	Sequence 787, App
c	30	33.6	10.0	449	16	US-10-424-599-51020	Sequence 51020, A
c	31	33.6	10.0	1210	13	US-10-027-632-216618	Sequence 216618,
c	32	33.6	10.0	1210	15	US-10-027-632-216618	Sequence 216618,
c	33	33.6	10.0	1770	13	US-10-027-632-259643	Sequence 259643,
c	34	33.6	10.0	1770	15	US-10-027-632-259643	Sequence 259643,
	35	33.6	10.0	17527	15	US-10-311-455-1406	Sequence 1406, Ap
	36	33.6	10.0	17527	16	US-10-240-454-28	Sequence 28, Appl
	37	33.4	9.9	528	17	US-10-021-323-8131	Sequence 8131, Ap
c	38	33.4	9.9	634	13	US-10-040-739-1384	Sequence 1384, Ap
	39	33.4	9.9	1829	17	US-10-437-963-6257	Sequence 6257, Ap
	40	33.4	9.9	3528	16	US-10-282-122A-9680	Sequence 9680, Ap
c	41	33.2	9.9	500	16	US-10-242-535A-54587	Sequence 54587, A
c	42	33.2	9.9	500	16	US-10-085-783A-54587	Sequence 54587, A
c	43	33.2	9.9	588	17	US-10-021-323-6082	Sequence 6082, Ap
c	44	33.2	9.9	1096	13	US-10-027-632-119180	Sequence 119180,
c	45	33.2	9.9	1096	13	US-10-027-632-119181	Sequence 119181,

ALIGNMENTS

RESULT 1

US-10-099-663-2

; Sequence 2, Application US/10099663

; Publication No. US20030177516A1

; GENERAL INFORMATION:

; APPLICANT: Avigenics, Inc

; TITLE OF INVENTION: Avian GUT-Specific Promoters

; FILE REFERENCE: A181

; CURRENT APPLICATION NUMBER: US/10/099,663

; CURRENT FILING DATE: 2002-03-14

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 336

; TYPE: DNA

; ORGANISM: Gallus gallus

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(336)

; OTHER INFORMATION: chick intestinal fatty acid binding protein promoter region

US-10-099-663-2

Query Match	100.0%	Score	336;	DB	15;	Length	336;
Best Local Similarity	100.0%	Pred. No.	9.4e-86;				
Matches	336;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	ATTATTATTTCATTAGATAGCGGGTTTTTACTCAAACTCAATAAGATGAACAGAAATG	60				
Db	1	ATTATTATTTCATTAGATAGCGGGTTTTTACTCAAACTCAATAAGATGAACAGAAATG	60				
Qy	61	AATCGGTTAGTCTGTTTATAAAGAGTATAAGATCACTATCATCATTTTGAGGCAA	120				
Db	61	AATCGGTTAGTCTGTTTATAAAGAGTATAAAGATCACTATCATCATTTTGAGGCAA	120				
Qy	121	TAAGGGAGGAGAGATTACGAAACAGTGTGCTTACAAAGTGGAACAACTAACTAAA	180				

Db 121 TAAGGGAGGAGAGATTTCAGCAACAGTGTGCTTACAGTGGAAAAACAAGTTAAACTAAA 180  
QY 181 GTGACCCCCCTCTTGCAAGATCAATGCCACAGTTGAGCTTTTAGCCAGCCACATCATCA 240  
Db 181 GTGACCCCCCTCTTGCAAGATCAATGCCACAGTTGAGCTTTTAGCCAGCCACATCATCA 240  
QY 241 TGTAAATTCCTTCTGTGATAGCCTGTTTCATAAATTCCTTTTGCAAGCTCTGCTACTTA 300  
Db 241 TGTAAATTCCTTCTGTGATAGCCTGTTTCATAAATTCCTTTTGCAAGCTCTGCTACTTA 300  
QY 301 CCAGAAGTCTGCCTACAGACAGAAAGATGGCATTTA 336  
Db 301 CCAGAAGTCTGCCTACAGACAGAAAGATGGCATTTA 336

RESULT 2  
US-10-099-663-1  
; Sequence 1, Application US/10099663  
; Publication No. US20030177516A1  
; GENERAL INFORMATION:  
; APPLICANT: Avigenics, Inc  
; TITLE OF INVENTION: Avian GUT-Specific Promoters  
; FILE REFERENCE: A181  
; CURRENT APPLICATION NUMBER: US/10/099,663  
; CURRENT FILING DATE: 2002-03-14  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2381  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)...(1626)  
; NAME/KEY: exon  
; LOCATION: (1627)...(1693)  
; NAME/KEY: Intron  
; LOCATION: (1694)...(2322)  
; NAME/KEY: exon  
; LOCATION: (2333)...(2381)  
US-10-099-663-1

Query Match 100.0%; Score 336; DB 15; Length 2381;  
Best Local Similarity 100.0%; Pred. No. 2.3e-85;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAATAAGATGAACAGAAATG 60  
Db 1301 ATTATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAATAAGATGAACAGAAATG 1360  
QY 61 AATGGGTTAGTGACTGTTTATAAAGAGATATAAAGATACATCATCATTTTGAGGCAA 120  
Db 1361 AATGGGTTAGTGACTGTTTATAAAGAGATATAAAGATACATCATCATTTTGAGGCAA 1420  
QY 121 TAAGGGAGGGAGAGATTTCAGCAACAGTGTGCTTACAAGTGGAAAAACAAGTTAAACTAAA 180  
Db 1421 TAAGGGAGGGAGAGATTTCAGCAACAGTGTGCTTACAAGTGGAAAAACAAGTTAAACTAAA 1480  
QY 181 GTGACCCCCCTCTTGCAAGATCAATGCCACAGTTGAGCTTTTAGCCAGCCACATCATCA 240  
Db 1481 GTGACCCCCCTCTTGCAAGATCAATGCCACAGTTGAGCTTTTAGCCAGCCACATCATCA 1540  
QY 241 TGTAAATTCCTTCTGTGATAGCCTGTTTCATAAATTCCTTTTGCAAGCTCTGCTACTTA 300  
Db 1541 TGTAAATTCCTTCTGTGATAGCCTGTTTCATAAATTCCTTTTGCAAGCTCTGCTACTTA 1600  
QY 301 CCAGAAGTCTGCCTACAGACAGAAAGATGGCATTTA 336  
Db 1601 CCAGAAGTCTGCCTACAGACAGAAAGATGGCATTTA 1636

RESULT 3

US-10-719-993-6867  
; Sequence 6867, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: PastSeq for Windows Version 4.0  
; SEQ ID NO 6867  
; LENGTH: 116327  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(116327)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-719-993-6867

Query Match 12.0%; Score 40.4; DB 18; Length 116327;  
Best Local Similarity 55.8%; Pred. No. 4.3;  
Matches 77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 197 ACAAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCT 256  
Db 70844 AAAAGATGATGCCCTCAACTAGTTTGACACAGTGATTTAGCATGCATAGAATTACCT 70903  
QY 257 GATAAGCCTGTTTCATAAATTCCTTTGCAAGCTCTGCTACTTACCAAGAGTCTGCCTAC 316  
Db 70904 GGTAGCTTTTAAAAAATTCATTTCCCAAGTCTTACTTCTAAAAAAGAGTCACTTTT 70963  
QY 317 AGACAGAAAGATGGCATT 334  
Db 70964 AGAAGAAAAAGAGTATT 70981

RESULT 4  
US-09-910-185-11/c  
; Sequence 11, Application US/09910185  
; Publication No. US20030083279A1  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-3 EXPRESSION  
; FILE REFERENCE: RTS-0258  
; CURRENT APPLICATION NUMBER: US/09/910,185  
; CURRENT FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 90  
; SEQ ID NO 11  
; LENGTH: 123526  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-910-185-11

Query Match 11.1%; Score 37.2; DB 10; Length 123526;  
Best Local Similarity 53.4%; Pred. No. 37;  
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 12 CATTAGATAGCCGGTTTTTTTACTACAACCTCAATAAGATGAACAGATGAATGGTTAGT 71  
Db 107264 CATTTCCTCAATGCTTTTTTAATTAATGCTACAGAAAATGCACAGAAAGAGGGTGCT 107205  
QY 72 GACTGTTTATAAAGAGAGTAAATAAGATACATCATCATTTGAGGCAATAAGGGAGGA 131  
Db 107204 GTGTACCCATTAAATAAATCAGATGAAGAAAAGAGAAAGGCAAGAGGGAGGA 107145  
QY 132 GAGATTGAGCAACAGTGTGCTTACA 157  
Db 107144 GGGAGAAAACCGAAAGCATCCATGCA 107119



```
RESULT 5
US-09-815-343-915
; Sequence 915, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(403)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-915

Query Match      10.9%; Score 36.6; DB 9; Length 403;
Best Local Similarity 50.6%; Pred. No. 4;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 26 TTTTCTACTACAACTCAAAATAAGATGAACAGAATGAATGGTTAGTGACTGTTTATAAAG 85
Db 178 TGTTTAAAGCTTAAAGAAAAAGAGCTGCAGAGTATTATTATAAACTGCTCTTTAGAAAAA 237

Qy 86 AAGAGTAATAAGATACATCATTTTGAGGCAATAAGGGAGGAGAGATTTCAGCAAAAC 145
Db 238 AACAAAGCAAGAACCATTTTGACCATATGAATGAAAAAGGGAAGAAAGTATTATAGAAG 297

Qy 146 AGTGTGCTTACAAGTGGAAAAACAAGTTAAACTAAAGTGACCCCCCT 191
Db 298 TTTGCTAGTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 343

RESULT 6
US-10-097-105-915
; Sequence 915, Application US/10097105
; Publication No. US20040037842A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Secrist, Heather
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504C1
; CURRENT APPLICATION NUMBER: US/10/097,105
; CURRENT FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 1562
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 308, 330, 334, 356, 359, 368, 369, 379, 383, 387, 389
; OTHER INFORMATION: n = A,T,C or G
US-10-097-105-915

Query Match      10.9%; Score 36.6; DB 16; Length 403;
Best Local Similarity 50.6%; Pred. No. 4;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
```

```
Qy 26 TTTTCTACTACAACTCAAAATAAGATGAACAGAATGAATGGTTAGTGACTGTTTATAAAG 85
Db 178 TGTTTAAAGCTTAAAGAAAAAGAGCTGCAGAGTATTATTATAAACTGCTCTTTAGAAAAA 237

Qy 86 AAGAGTAATAAGATACATCATTTTGAGGCAATAAGGGAGGAGAGATTTCAGCAAAAC 145
Db 238 AACAAAGCAAGAACCATTTTGACCATATGAATGAAAAAGGGAAGAAAGTATTATAGAAG 297

Qy 146 AGTGTGCTTACAAGTGGAAAAACAAGTTAAACTAAAGTGACCCCCCT 191
Db 298 TTTGCTAGTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 343

RESULT 7
US-10-221-613-217
; Sequence 217, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 217
; LENGTH: 8648
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-217

Query Match      10.8%; Score 36.2; DB 16; Length 8648;
Best Local Similarity 57.5%; Pred. No. 21;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 3 TATTATTTTCATTAGATAGCCGGTTTTTACTACAACTCAAAATAAGATGAACAGAATGAA 62
Db 940 TTTTGGTTTTATTACGTGGTTAGTTATTATTAGTATGATTATGATAAGTTTGTGATAT 999

Qy 63 TGGGTTAGTCAGCTTTTATAAAGAGAGTAATAAAGATACATCATCTTGA 115
Db 1000 TTGGGAAGTCAGTATTTTATGTGAGAGTATAGATTATTTTGTATTGA 1052

RESULT 8
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MH
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match      10.5%; Score 35.4; DB 15; Length 3673778;
Best Local Similarity 53.2%; Pred. No. 4.7e+02;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 38 ACTCAATAAGATGAACAGATGAATGGTTAGTGACTGTTTATAAGAGAGAGTAATAAA 97
DB 1311399 ATTTTAATTAATTATTAATTAATAGGTAAATAATAGGATAAAAAATAAGAGAGAAA 1311518

QY 98 GATACTATCATCTTTGAGGCAATTAAGGGAGGAGAGATTGAGCAACAGTGTGCTTACA 157
DB 1311459 GATACGGTTATATTGTATGTTAGTTATGTTAATTAATTAGAAAAAATTTGAGTGATT 1311518

QY 158 AGTGGAAAAACAGTTAAACTA 178
DB 1311519 ATTTTAAAAATTAGATAAATA 1311539
```

```

RESULT 9
US-10-242-535A-47091/c
; Sequence 47091, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 47091
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-47091
```

```

Query Match      10.5%; Score 35.2; DB 16; Length 479;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 236 CATCATGTAATTAATGCTTTCTCTGATAGCCCTGTTTCATAAATTCCTTTGCAAGCTCTGCT 295
DB 377 CACCATGCCAACTGCTTTCAAGTTTATTATTTCAGAAATACTTTTTGCAGGCTGTAGC 318

QY 296 ACTTACAGAAAGTCTGCCTACAGACAGA 323
DB 317 TTCTACAAAAAGTAATTCCTTCAGATAGA 290
```

```

RESULT 10
US-10-085-783A-47091/c
; Sequence 47091, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
```

```

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 47091
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-47091
```

```

Query Match      10.5%; Score 35.2; DB 16; Length 479;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 236 CATCATGTAATTAATGCTTTCTCTGATAGCCCTGTTTCATAAATTCCTTTGCAAGCTCTGCT 295
DB 377 CACCATGCCAACTGCTTTCAAGTTTATTATTTCAGAAATACTTTTTGCAGGCTGTAGC 318

QY 296 ACTTACAGAAAGTCTGCCTACAGACAGA 323
DB 317 TTCTACAAAAAGTAATTCCTTCAGATAGA 290
```

```

RESULT 11
US-10-311-455-2115
; Sequence 2115, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-15
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2115
; LENGTH: 8622
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2115
```

```

Query Match      10.5%; Score 35.2; DB 15; Length 8622;
Best Local Similarity 45.1%; Pred. No. 40;
Matches 130; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1 ATTTATTATTTTCATTAGATAGCCGGTTTTTTTACTACAACTCAATAATAGATGAACAGAAATG 60
DB 2566 ATGTTTATTTTGGAAAAAGAAATGTGTAATTTTTTAAAAAGGTTAAATATGTATTATATG 2625

QY 61 AATGGTTAGTGACTGTTTATATAAGAGAGTAATAAGATACATCATCTTTTCAGGCAA 120
DB 2626 ATTTAGTAAATTTATTTTAGATATATAATTAAGAAATTAATAATATAGATTTTATAAAA 2685

QY 121 TAAGGGAGGAGAGATTACAGCAACAGTGTGCTTTACAAGTGGAAAAACAAAGTTAAACTAAA 180
```

Db 2686 AAATTTGTATATAAATGTTTATATAGTATATATTTTAAATAGTTTATAAAGTAGAAAAATAA 2745  
QY 181 GTGACCCCTCCCTTGCAAGATCAATGCGACAGCTTGAGCTTTAGCCAGCCACATCATCA 240  
Db 2746 TTAATGTTTATCGGTGATGATGAATAAATAAATATGTTGTGATATGTTTATATATA 2805  
QY 241 TGTAAATTTGCTTCTCTGATAGCTGTTTCATAAATCTCTTTTGCAAG 288  
Db 2806 GATATTATTGCTTTATGAAAAATATATATAAATAATATTATGTTAAATG 2853

RESULT 12  
US-10-311-455-1146/c  
; Sequence 1146, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation  
; TITLE OF INVENTION: cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311.455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043926.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1146  
; LENGTH: 15714  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1146

Query Match 10.5%; Score 35.2; DB 15; Length 15714;  
Best Local Similarity 50.6%; Pred. No. 53; Mismatches 0; Indels 0; Gaps 0;  
Matches 85; Conservative 0;  
QY 154 TACAAGTGGAAAAACAAGTTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACA 213  
Db 10154 TAAAAATAAAAAACAATATCAAAAAAATCTCAAAATCTTCCAAAAAATATACAAA 10095  
QY 214 GTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTTGATAAGCTGTTCTATAA 273  
Db 10094 ATCTTACCAATCTATCTCTCAAAACCACATACAAATTTACCTTCTTTACAATACATACATAA 10035  
QY 274 ATTCTCTTTGCAAGCTCTGCTACTTACCAGAGTCTGCGCTACAGACA 321  
Db 10034 AATACTCTAAAAATCAAAACTAATATTATTAATAAACTTCTTAACCAAAA 9987

RESULT 13  
US-10-433-793-88/c  
; Sequence 88, Application US/10433793  
; Publication No. US20040142334A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/433.793  
; CURRENT FILING DATE: 2003-06-06  
; NUMBER OF SEQ ID NOS: 212  
; SEQ ID NO 88  
; LENGTH: 15714  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-433-793-88  
Query Match 10.5%; Score 35.2; DB 17; Length 15714;  
Best Local Similarity 50.8%; Pred. No. 53; Mismatches 0; Indels 0; Gaps 0;  
Matches 85; Conservative 0;  
QY 154 TACAAGTGGAAAAACAAGTTAAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACA 213  
Db 10154 TAAAAATAAAAAACAATATCAAAAAAATCTTCCAAAAAATATACAAA 10095  
QY 214 GTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTTGATAAGCTGTTCTATAA 273  
Db 10094 ATCTTACCAATCTATCTCTCAAAACCACATACAAATTTACCTTCTTTACAATACATACATAA 10035  
QY 274 ATTCTCTTTGCAAGCTCTGCTACTTACCAGAGTCTGCGCTACAGACA 321  
Db 10034 AATACTCTAAAAATCAAAACTAATATTATTAATAAACTTCTTAACCAAAA 9987

RESULT 14  
US-10-087-192-502/c  
; Sequence 502, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087.192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 502  
; LENGTH: 134738  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(134738)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-502

Query Match 10.4%; Score 35; DB 13; Length 134738;  
Best Local Similarity 53.5%; Pred. No. 1.6e+02;  
Matches 99; Conservative 0; Mismatches 80; Indels 6; Gaps 1;  
QY 2 TTATTTATTTTCATTAGATAGCCGGTTTTTTTACTACAACCTCAATAAGATGAACAGATGA 61  
Db 25555 TTTTATGTACTTACCAGATTTTAGTCTTTTGTATTAATCTACTATATAAATATGAACAAATCA 25496  
QY 62 ATGGG-----TTAGTGACTGTTTATAAAGAGAGTAATAAGATACATCATCATTTGA 115  
Db 25495 AAGGGGCATCTTGGCTGTATGTTTTTTAACTGCATAATAGACAGCCTATATTTATTTAA 25436  
QY 116 GGCAATAAGGGAGGAGAGATTCAGCAACACAGTGTCTTACAAAGTGGAACAAAGTTAAA 175  
Db 25435 GTTAAAAAGGGCGGCAAGAATAGTAAGAAAGAGGCTTTTAAATGACATCAGAAAAGGTATA 25376  
QY 176 CTAAAA 180  
Db 25375 TAAAAA 25371

RESULT 15  
US-09-854-883-243/c  
; Sequence 243, Application US/09854883  
; Patent No. US20020055479A1  
; GENERAL INFORMATION:

```
; APPLICANT: Lex M. Cowbert
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PPIB EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/09/854,883
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629,644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487,368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO 243
; LENGTH: 75899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-854-883-243
```

```
Query Match      10.4%; Score 34.8; DB 9; Length 75899;
Best Local Similarity 60.6%; Pred. No. 1.4e+02;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy      26  TTTTCTACTCAACTCAAATAGATGAACAGAAATGAATGGTTAGTGCCTGTTTATAAG 85
Db      27384 TTGTTTCTACGTAACATGAATATAGCATAAAAAAGTAAGTATTCTTAAAC 27325

Qy      86  AAGAGTAATAAGATCTATCATCATTTTGAGGCA 119
Db      27324 AAGAGGTAATAAGATATAATCAAGACTTGGACA 27291
```

Search completed: January 14, 2005, 19:33:52  
Job time : 208.381 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	41.8	12.4	12.4	500	1	AU088319	AU088319 AU088319
2	41.4	12.3	12.3	444	5	BW139785	BW139785 BW139785
C 3	41.4	12.3	12.3	504	5	BW269615	BW269615 BW269615
4	41.4	12.3	12.3	574	1	AV680413	AV680413 AV680413
5	41.4	12.3	12.3	720	5	BW139668	BW139668 BW139668
C 6	41.4	12.3	12.3	720	5	BW186283	BW186283 BW186283
7	41.4	12.3	12.3	735	5	BW140184	BW140184 BW140184
8	41.4	12.3	12.3	732	5	BW127139	BW127139 BW127139
9	41.4	12.3	12.3	759	1	AV877102	AV877102 AV877102
10	41.4	12.3	12.3	767	5	BW082303	BW082303 BW082303
11	41.4	12.3	12.3	774	5	BW141206	BW141206 BW141206
12	41	12.2	12.2	598	6	CF250818	CF250818 esa017_e0
C 13	40.6	12.1	12.1	594	8	AZ305625	AZ305625 1M0006E20
C 14	40.6	12.1	12.1	732	9	AG596239	AG596239 Mus muscu
C 15	40.6	12.1	12.1	781	9	AG470849	AG470849 Mus muscu
16	40.4	12.0	12.0	554	5	BW009156	BW009156 BW009156
C 17	40.2	12.0	12.0	522	8	AZ431375	AZ431375 1M0216J16
C 18	40	11.9	11.9	879	8	BH137539	BH137539 ENTOM83TR
C 19	40	11.9	11.9	943	8	BH147340	BH147340 ENTOM834TR
C 20	39.8	11.8	11.8	472	5	BW093545	BW093545 BW093545
C 21	39.8	11.8	11.8	507	5	BW234156	BW234156 BW234156
22	39.8	11.8	11.8	688	5	BW133492	BW133492 BW133492
C 23	39.8	11.8	11.8	760	5	BW399296	BW399296 BW399296
C 24	39.2	11.7	11.7	1101	9	CNS00BQ6	AL069511 Drosophi1



REFERENCE 1 (bases 1 to 574)  
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.  
TITLE Expressed genes in Ciona intestinalis  
JOURNAL Unpublished (2000)  
COMMENT Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES  
source 1. .574

Location/Qualifiers  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clones="rcitb12j8"  
/tissue\_type="whole animal"  
/dev\_stage="tailbud"  
/clone\_lib="Nori Satoh unpublished cDNA library"

## ORIGIN

Query Match 12.3%; Score 41.4; DB 1; Length 574;  
Best Local Similarity 46.4%; Pred. No. 1.2;  
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
  
QY 35 ACAACTCAATAAGATGAACAGAGATGAATGGTTAGTGACTGTTTATAAAGAAGAGTAAT 94  
DB 36 ATAACTGTATTTCAGAAAGCAGAACATATGTGTTCCAAACGGTCAACAATAAGGTAAT 95  
QY 95 AAAGATACTATCATCATTTGAGGCAATAAGGGAGGAGAGATTACGACAAACAGTGTGCTT 154  
DB 96 GACAGTTCCCAATACCTTTCTTGTAATAATATGACCCCTGCTTCATTTTACATTGACCA 155  
QY 155 ACAAGTGGAAACAAGTTAACTAAAGTACCCCTCTTGACAAAGATCAATGCCACAG 214  
DB 156 CGAAATGGAATAGGTGTTAAATAATAGTAAACATTAATAAATGTTACCGAAATGACCCCTG 215  
QY 215 TTGAGCTTTAGCAGCCACATCATCATGTTAAATGCTTCTGATAGAGCTGTTTCATAAA 274  
DB 216 CTAGCATCTGCCATTGCTGTATTTAATCTGTGTCTTCTTGACAAATAGTTGCAAC 275  
QY 275 TTCTTTTGCAGAGCTCTGCTACTTACCAGAGTCTGCCTACAGACAGAAA 325  
DB 276 AGCTTTCTGTAATACTCCACACTTGCAATTCAGTTCTGCTCTGCTCTCTAA 326

RESULT 5  
BW135968 720 bp mRNA linear EST 02-NOV-2002  
LOCUS BW135968 Nori Satoh unpublished cDNA library, gastrula and neurula  
DEFINITION Ciona intestinalis cDNA clone rcign042b07 3', mRNA sequence.  
ACCESSION BW135968  
VERSION BW135968.1 GI:24492367  
KEYWORDS EST.  
SOURCE Ciona intestinalis  
ORGANISM Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 720)  
Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.  
TITLE Expressed genes in Ciona intestinalis (2002c)  
JOURNAL Unpublished (2002)  
COMMENT Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES  
source 1. .720

Location/Qualifiers  
/organism="Ciona intestinalis"

/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="rcign042b07"  
/tissue\_type="whole body"  
/dev\_stage="gastrula and neurula"  
/clone\_lib="Nori Satoh unpublished cDNA library, gastrula and neurula"

## ORIGIN

Query Match 12.3%; Score 41.4; DB 5; Length 720;  
Best Local Similarity 46.4%; Pred. No. 1.2;  
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
  
QY 35 ACAACTCAATAAGATGAACAGAGATGAATGGTTAGTGACTGTTTATAAAGAAGAGTAAT 94  
DB 24 ATAACTGTATTTCAGAAAGCAGAACATATGTGTTCCAAACGGTCAACAATAAGGTAAT 83  
QY 95 AAAGATACTATCATCATTTGAGGCAATAAGGGAGGAGAGATTACGACAAACAGTGTGCTT 154  
DB 84 GACAGTTCCCAATACCTTTCTTGTAATAATATGACCCCTGCTTCATTTTACATTGACCA 143  
QY 155 ACAAGTGGAAACAAGTTAACTAAAGTACCCCTCTTGACAAAGATCAATGCCACAG 214  
DB 144 CGAAATGGAATAGGTGTTAAATAATAGTAAACATTAATAAATGTTACCGAAATGACCCCTG 203  
QY 215 TTGAGCTTTAGCAGCCACATCATCATGTTAAATGCTTCTGATAGAGCTGTTTCATAAA 274  
DB 204 CTAGCATCTGCCATTGCTGTATTTAATCTGTGTCTTCTTGACAAATAGTTGCAAC 263  
QY 275 TTCTTTTGCAGAGCTCTGCTACTTACCAGAGTCTGCCTACAGACAGAAA 325  
DB 264 AGCTTTCTGTAATACTCCACACTTGCAATTCAGTTCTGCTCTGCTCTCTAA 314

## RESULT 6

BW186283/c 725 bp mRNA linear EST 05-NOV-2002  
LOCUS BW186283 Nori Satoh unpublished cDNA library, heart Ciona  
DEFINITION Ciona intestinalis cDNA clone rciht035k14 3', mRNA sequence.  
ACCESSION BW186283  
VERSION BW186283.1 GI:24576615  
KEYWORDS EST.  
SOURCE Ciona intestinalis  
ORGANISM Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 725)  
Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.  
TITLE Expressed genes in Ciona intestinalis (2002c)  
JOURNAL Unpublished (2002)  
COMMENT Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES  
source 1. .725

Location/Qualifiers  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="rciht035k14"  
/tissue\_type="heart"  
/clone\_lib="Nori Satoh unpublished cDNA library, heart"

## ORIGIN

Query Match 12.3%; Score 41.4; DB 5; Length 725;  
Best Local Similarity 46.4%; Pred. No. 1.2;  
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
  
QY 35 ACAACTCAATAAGATGAACAGAGATGAATGGTTAGTGACTGTTTATAAAGAAGAGTAAT 94

```

Db      304 ATAACTGTTATTCAGAAAGCAGAAACATATGTGTTCAAAACGGTCAACAAATAAGGTAAAT 245
QY      95 AAGATATCATCATCATTTTGGAGCAATAAGGGAGGAGAGATTTCAGCAAAACAGTGTGTT 154
Db      244 GACAGTTCGAATACCTTTTCTGTAAAAAATATGACCCCTGCTTCATTTTACATTGCACA 185
QY      155 ACAAGTGGAAACAAAGTTAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACAG 214
Db      184 CGAAATGGAAATAGGTGTAAAAATAGTAAACATTAATAAATGTACCGAAATGCACCCCTG 125
QY      215 TTGAGCTTTAGCCAGCACATCATCATGTAAATTCGTTTCCGATAAGCCTGTTCAATAA 274
Db      124 CTAAGCATCTGCCATTGCTGTATTTAACTTCGTGTCTTCTTGACAACATAGTTGCAAC 65
QY      275 TTCTCTTTTCAAGCTCTGCTACTTACCAAGAAGTCTGCCCTACAGACAGAAA 325
Db      64 AGCTTCTGTAAATCTCCACCATTGCATTTCAGTTCTGCTTGCCTCTAA 14

RESULT 7
LOCUS   BW140184
DEFINITION   BW140184 Nori Satoh unpublished cDNA library, linear EST 03-NOV-2002
              Ciona intestinalis cDNA clone rcign055e20 3', mRNA sequence.
ACCESSION   BW140184
VERSION     BW140184.1 GI:24497206
KEYWORDS    EST.
SOURCE      Ciona intestinalis
ORGANISM    Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Clonidae; Ciona.

REFERENCE   1 (bases 1 to 730)
AUTHORS    Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)
JOURNAL    Unpublished (2002)
COMMENT    Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES             source
     1..730
     /organism="Ciona intestinalis"
     /mol_type="mRNA"
     /db_xref="taxon:7719"
     /clone="rcign055e20"
     /tissue_type="whole body"
     /dev_stage="gastrula and neurula"
     /clone_lib="Nori Satoh unpublished cDNA library, gastrula
     and neurula"

ORIGIN

Query Match      12.3%; Score 41.4; DB 5; Length 730;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY      35 ACAACTCAATAAGATGAACAGAAATGAATGGGTAGTACTGTTTATAAAGAGAGTAAT 94
Db      40 ATAAGTGTATTCAGAAAGCAGAAACATATGTGTTCACACGGTCAACAAATAAGGTAAAT 99
QY      95 AAAGATACATCATCATTTGAGCAATAAGGGAGGAGAGATTTCAGCAAAACAGTGTGTT 154
Db      100 GACAGTTCATACCTTTTCTGTAAAAAATATGACCCCTGCTTCATTTTACATTGCACA 159
QY      155 ACAAGTGGAAACAAAGTTAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACAG 214
Db      160 CGAAATGGAAATAGGTGTAAAAATAGTAAACATTAATAAATGTACCGAAATGCACCCCTG 219
QY      215 TTGAGCTTTAGCCAGCACATCATCATGTAAATTCGTTTCCGATAAGCCTGTTCAATAA 274
Db      220 CTAAGCATCTGCCATTGCTGTATTTAACTTCGTGTCTTCTTGACAACATAGTTGCAAC 279

```

```

QY      275 TTCTCTTTGCAAGCTCTGCTACTTACCAAGAAGTCTGCCCTACAGACAGAAA 325
Db      280 AGCTTCTGTAAATCTCCACCATTGCATTTCAGTTCTGCTTGCCTCTAA 330

RESULT 8
LOCUS   BW127139
DEFINITION   BW127139 Nori Satoh unpublished cDNA library, linear EST 03-NOV-2002
              Ciona intestinalis cDNA clone rcign013j16 3', mRNA sequence.
ACCESSION   BW127139
VERSION     BW127139.1 GI:24483538
KEYWORDS    EST.
SOURCE      Ciona intestinalis
ORGANISM    Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
              Phlebobranchia; Clonidae; Ciona.

REFERENCE   1 (bases 1 to 732)
AUTHORS    Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)
JOURNAL    Unpublished (2002)
COMMENT    Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES             source
     1..732
     /organism="Ciona intestinalis"
     /mol_type="mRNA"
     /db_xref="taxon:7719"
     /clone="rcign013j16"
     /tissue_type="whole body"
     /dev_stage="gastrula and neurula"
     /clone_lib="Nori Satoh unpublished cDNA library, gastrula
     and neurula"

ORIGIN

Query Match      12.3%; Score 41.4; DB 5; Length 732;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY      35 ACAACTCAATAAGATGAACAGAAATGAATGGGTAGTACTGTTTATAAAGAGAGTAAT 94
Db      20 ATAAGTGTATTCAGAAAGCAGAAACATATGTGTTCACACGGTCAACAAATAAGGTAAAT 79
QY      95 AAAGATACATCATCATTTGAGCAATAAGGGAGGAGAGATTTCAGCAAAACAGTGTGCTT 154
Db      80 GACAGTTCATACCTTTTCTGTAAAAAATATGACCCCTGCTTCATTTTACATTGCACA 139
QY      155 ACAAGTGGAAACAAAGTTAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACAG 214
Db      140 CGAAATGGAAATAGGTGTAAAAATAGTAAACATTAATAAATGTACCGAAATGCACCCCTG 199
QY      215 TTGAGCTTTAGCCAGCACATCATCATGTAAATTCGTTTCCGATAAGCCTGTTCAATAA 274
Db      200 CTAAGCATCTGCCATTGCTGTATTTAAATTCGTGTCTTCTTGACAACATAGTTGCAAC 259
QY      275 TTCTCTTTGCAAGCTCTGCTACTTACCAAGAAGTCTGCCCTACAGACAGAAA 325
Db      260 AGCTTCTGTAAATCTCCACCATTGCATTTCAGTTCTGCTTGCCTCTAA 310

RESULT 9
LOCUS   AV877102
DEFINITION   AV877102 Nori Satoh unpublished cDNA library, linear EST 08-NOV-2001
              intestinalis cDNA clone rcitb31k05 3', mRNA sequence.
ACCESSION   AV877102
VERSION     AV877102.1 GI:16864626

```



```
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 759)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES source
1..759
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcitb31k05"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cDNA library, tailbud
embryo"
ORIGIN
Query Match 12.3%; Score 41.4; DB 1; Length 759;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
Qy 35 ACRACTCAATAAGATGAACAGAGTAATGGGTTAGTGACTGTTTATAAGAGAGTAAT 94
Db 453 ATAACTGTATTACGAAGCAGAAACATATGTGTTCCAAACGGTCAACAAATAAGTAAT 512
Qy 95 AAAGATACTATCATCTTTGAGGCAATAAGGGAGGAGAGATTGAGCAACAGTGCTT 154
Db 513 GACAGTTCCAAATACCTTTCTGTGTAATAATATGACCTGCTTCATTTTACATTGCACCA 572
Qy 155 ACAAGTGGAAACAAGTTAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACAG 214
Db 573 CGAAATGGAATAGGTGTGTTAAATAATAGTAAACATTTAAAAATTTGTTACCGAAATGCACCCCTG 632
Qy 215 TTGAGCTTTAGCCAGCCACATCATCATGTAATGCTTTCTCTGATAGAGCTGTTTATAAA 274
Db 633 CTAAGCATCTGCCATTGCTGTATTTAACTCTGTGTGTTCTTGACAAACATAGTTGCAAC 692
Qy 275 TTCTCTTTGCAAGCTCTGCTACTTACCAAGAGTCTGCCTACAGACAGAAA 325
Db 693 AGCTTTCTGTAATAATCTCACACATTGCATTGCAATTCAGTTCTGCTTCTGCCTCTAA 743
RESULT 10
BW082303 767 bp mRNA linear EST 22-OCT-2002
LOCUS BW082303 Nori Satoh unpublished cDNA library, egg Ciona
DEFINITION intestinalis cDNA clone rcieg088c22 3', mRNA sequence.
ACCESSION BW082303.1 GI:24257583
VERSION BW082303.1
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 767)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES source
1..774
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcign058h04"
/tissue_type="whole body"
/dev_stage="gastrula and neurula"
/clone_lib="Nori Satoh unpublished cDNA library, gastrula
and neurula"
```

```
FEATURES source
1..767
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcieg088c22"
/tissue_type="whole animal"
/dev_stage="egg"
/clone_lib="Nori Satoh unpublished cDNA library, egg"
ORIGIN
Query Match 12.3%; Score 41.4; DB 5; Length 767;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
Qy 35 ACRACTCAATAAGATGAACAGAGTAATGGGTTAGTGACTGTTTATAAGAGAGTAAT 94
Db 27 ATAACTGTATTACGAAGCAGAAACATATGTGTTCCAAACGGTCAACAAATAAGTAAT 86
Qy 95 AAAGATACTATCATCTTTGAGGCAATAAGGGAGGAGAGATTGAGCAACAGTGCTT 154
Db 87 GACAGTTCCAAATACCTTTCTGTGTAATAATATGACCTGCTTCATTTTACATTGCACCA 146
Qy 155 ACAAGTGGAAACAAGTTAACTAAAGTGACCCCTCTCTTGACAAGATCAATGCCACAG 214
Db 147 CGAAATGGAATAGGTGTGTTAAATAATAGTAAACATTTAAAAATTTGTTACCGAAATGCACCCCTG 206
Qy 215 TTGAGCTTTAGCCAGCCACATCATCATGTAATGCTTTCTCTGATAGAGCTGTTTATAAA 274
Db 207 CTAAGCATCTGCCATTGCTGTATTTAACTCTGTGTGTTCTTGACAAACATAGTTGCAAC 266
Qy 275 TTCTCTTTGCAAGCTCTGCTACTTACCAAGAGTCTGCCTACAGACAGAAA 325
Db 267 AGCTTTCTGTAATAATCTCACACATTGCATTGCAATTCAGTTCTGCTTCTGCCTCTAA 317
RESULT 11
BW141206 774 bp mRNA linear EST 03-NOV-2002
LOCUS BW141206 Nori Satoh unpublished cDNA library, gastrula and neurula
DEFINITION Ciona intestinalis cDNA clone rcign058h04 3', mRNA sequence.
ACCESSION BW141206
VERSION BW141206.1 GI:24498431
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 774)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES source
1..774
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcign058h04"
/tissue_type="whole body"
/dev_stage="gastrula and neurula"
/clone_lib="Nori Satoh unpublished cDNA library, gastrula
and neurula"
```

```
Query Match      12.3%; Score 41.4; DB 5; Length 774;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 35 ACAACTCAATTAAGATGAACAGATGAATGGGTAGTACTGTTTATTAAGAAGTAAAT 94
Db 20 ATAACGTATTACGAAAGCAGACATATGTGTTCCAAACGGTCAACAAATAAGGTAAAT 79
QY 95 AAAGATACTATCATCTTTGAGGCAATAAGGAGGAGAGATTACGACAAACAGTGTGCTT 154
Db 80 GACAGTCCAAATACCTTTTCTGTAAAAAATATGACCCCTGCTTCATTTTACATTGCACCA 139
QY 155 ACAAGTGGAAAAACAAGTTAAACTTAAGTACCCCTCTTGCACAGATCAATGCCACAG 214
Db 140 CGAATGGATAGGTGTTAAATATGAACATTAACAAATTTACCGAAATGCACCCCTG 199
QY 215 TTGAGCTTTAGCCGCCACATCATCATGATAAATTGCTTCTGATAAGCCTGTTTCAAAA 274
Db 200 CTAAGCATCTGCCTATTGCTGTAATTAATCTGTGTGTTCTTGCACACATAGTTGCAAC 259
QY 275 TTCTCTTTGCAAGCTCTGCTACTTACCAGAAAGTCTGCTACACAGACAGAAA 325
Db 260 AGCTTTCTGTAATCTCCACCATTGCATTTCAGTTCTGCTTCTGCTCTAA 310

RESULT 12
LOCUS CF250818
DEFINITION esb017_e05 Eimeria tenella-infected caecal tonsil Gallus gallus
cDNA, mRNA sequence.
ACCESSION CF250818
VERSION CF250818.1 GI:33484073
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 598)
Witzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,
Chausse,A.M. and Zoorob,R.
A collection of chicken ESTs from activated immune cells
Unpublished (2003)
Contact: Zoorob R
UPR 1983
CNRS
7 rue Guy Moquet, Bp 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoorob@vjf.cnrs.fr.
Location/Qualifiers
1..598
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone_lib="Eimeria tenella-infected caecal tonsil"
/notes="Organ: Caecal tonsil; Vector: pRipLEX2"

ORIGIN
Query Match      12.2%; Score 41; DB 6; Length 598;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 ACTTACCAGAGTCTGCCTACACAGACAGAAAGATGGCATTTA 336
Db 3 ACTTACCAGAGTCTGCCTACACAGACAGAAAGATGGCATTTA 43

RESULT 13
LOCUS AZ305625/c
DEFINITION 1M0006F20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
cDNA, mRNA sequence.
ACCESSION AZ305625
VERSION AZ305625.1 GI:10342825
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 594)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islan,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0006 row: F column: 20
Seq primer: CGTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 594.
Location/Qualifiers
1..594
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0006F20"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 [gil4732114|gb|AF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

Db	388	TAGAGTGGGGGAAAGAGAAAGTTAAAAAAGAA	350
RESULT 14			
AG596239/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy	22	CCGGTGTCTTAC	81
Db	656	CTGATTATTTCATCAAGAGAAAGAAACAAAGGAATNAAGGAGAAAGTCAGAAAAAGAA	597
Qy	82	AAAGAGAGTAAATAAGATATCATCTATTTCAGGCAATAAGGGAGGAGAGATTTCAGC	141
Db	596	AAAGAGATGAAGGAAGAAAGAAAGGAAGTAAGGAAGGAGGAGCGGAGGAGG	537
Qy	142	AAACAGTGTCTTACAGTGGAAAAACAAGTTAAACTAAA	180
Db	536	TAGAGTGGGGGAAAGAGAAAGTTAAAAAAGAA	498

RESULT 15	DNA	linear	GSS 04-JUN-2004
AG470649/c			
LOCUS	781 bp		
AG470649			

**THIS PAGE IS BLANK**